Reexamining Computational Support for Intelligence Analysis:
A Functional Design for a Future Capability

Drs. James Llinas (Principal Investigator) and Galina Rogova
Center for Multisource Information Fusion (CMIF)
State University of New York at Buffalo
llinas, rogova@buffalo.edu
and
Messrs Kevin Barry (Pgm Mgr), Rachel Hingst, Peter Gerken, Dr. Alicia Ruvinsky
Lockheed Advanced Technology Laboratories (ATL)
Cherry Hill, New Jersey
kevin.barry@lmco.com

1. Introduction

This technical report is the Final Deliverable for Grant No.: N00244-15-1-0051 from the Naval Postgraduate School (NPS) with funding via the NAVSUP Fleet Logistics Center San Diego, on the topic of "Advancing Human-Machine Symbiosis Using Hybrid Methods for Collaborative Problem-Solving". This effort was a basic and applied research study program involving thorough reviews and assessments of applicable state of the art research and methods, and the development of a detailed functional design for an advanced-capability intelligence analysis computer-based support system. Our team comprised faculty from the State University of New York at Buffalo (aka University at Buffalo (UB)) and technical staff from the Advanced Technology Laboratories of Lockheed Corporation (LMCO/ATL) as a subcontractor partner. The technical areas studied involved principles of argumentation including especially computational support aspects for enabling argumentation-based analysis, story-based analysis, uncertainty aspects of analysis in an open-world environment, human-machine symbiosis, hard and soft information fusion, and computational methods for narrative development. Project thinking in regard to intelligence based analysis was vetted by a number of thrusts to include a visit to the National Air and Space Intelligence Center (NASIC) at Wright-Patterson Air Force Base in Dayton Ohio where our team met with staff from the Advanced Analytics Cell (AAC); broadly speaking that visit confirmed the general validity of our goals and objectives for this work. In addition, discussions were held with government staff and Subject Matter Experts (SME's) who were experienced intelligence analysts from Army intelligence community regarding the general issue of improving rigor in intelligence analysis, a main thrust of the Army Training and Doctrine Command (TRADOC). Our effort cumulates in a top-level functional design of a notional prototype capability for providing computational support to a hybrid argumentation plus story-based analysis capability. This research was further vetted in the presentation and writing of a paper on the project at the 13th International Conference on Distributed Computing and Artificial Intelligence (DCAI'16) in Seville, Spain in June of 2016 [1].

1.1 Acknowledgement

This publication results from research supported by the Naval Postgraduate School

Assistance Grant No. ***N00244-15-1-0051*** *awarded by the NAVSUP Fleet Logistics Center San Diego (NAVSUP FLC San Diego). The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.*

2. Motivation

Analysis Tool Suites (ATS's) such as Analyst's Notebook [2], Analyst's Workspace [3], Sentinel Visualizer [4], and Palantir Government [5], Entity Workspace [6], and Jigsaw [7], among others are examples of modern intelligence analysis frameworks. A major point for sensibly all these tool-suites are that they start by focusing on the entity level within the environments of interest. None overtly discuss computational support to inter-entity association and attribute/relation fusion. That is, most if not all are single-source-based as regards entity streams, with the tools doing varying degrees of automated link analysis among bounded entity-pairs toward realization of "data fusion" albeit with rather limited rigor. Further, most also assume that any preprocessing that provides entity extraction yields correct results. This framework of tool products provides the basis for identifying and visualizing relational connections between entities, but these connections are largely if not exclusively performed in the mind of the analyst. In most cases, nothing is done in the way of computational support to dealing with entity or relational uncertainties. The primary function of most of these ATS's is relational link discovery to discern inter-entity relations of bounded extent (in graph science terminology, usually single-hop or limited-hop relations), achieved with quite limited analytical formality regarding issues of uncertainty, inter-data and/or inter-entity associability, and of relational complexities. Thus, deeper and broader analysis of entity and relational connectedness is left for the human analyst. This is especially true in regard to the assembly of typical final desired analysis products in the form of stories or narratives; said otherwise, there is very limited technical support for synthesis or fusion of hypotheses into the larger context of situational understanding. By and large, these tools try to support the Sensemaking (SM) or schema-development loop of SM [8, 9], but either have no algorithmic or technological-process support or provide quite-limited automated support to these higher goals; these assessments are summarized in the review paper of [10].

Thus we perceive a need first for a processing/reasoning paradigm that can provide the framework for a more holistic, systemic based approach to intelligence analysis. As sensibly all critiques about intelligence analysis as well as the analysis requirements stated in field manuals describe that the main product that an analyst is driving toward is a narrative type description of some world condition/situation, we set this goal for our project as well. So, primarily we are seeking to study ways that discrete, single-theme hypotheses can be synthesized or fused into a more holistic and semantic construct in the form of a story or narrative. We also want to exploit our team's unique prior history in associating and fusing so-called hard (sensor) and soft (textual, semantic) information, as many intelligence analysis environments have such disparate data streams as input. (We note that virtually all the work in the areas we studied here only involve soft or textual type inputs.) We believe that the functional design produced here provides a basis for a next step involving research prototype development, and because of this we have also studied ways to test and evaluate such a prototype.

3. Symbiosis

Symbiosis is a term that can be mean a pretty broad range of concepts. In its most fundamental sense the term evolves from the Greek for "living together" (wiki, whatever), usually meaning that the two participants have a mutually beneficial relationship and interaction. As the term was originally conjured to describe biological type relationships between living entities, some complexity arises when one participant is the computer, a machine, and the other a human being. While some early conceptualizations of artificial intelligence spoke of symbiosis between some type of machine and humans, the most frequently cited reference to human-computer symbiosis is to Joseph Licklider's paper of 1960 titled "Man-Computer Symbiosis" [11]. Licklider's framing of the symbiosis idea was in the spirit of an augmentation of human cognition resulting from whatever interactions occurred between the human and the computer [12]. A few years later Engelbart publishes a somewhat different view asserting that the symbiosis yields amplification in human intellect [13]. These notions are related and we will not drift into the subtleties but cognition seems to relate to the process of thinking and reasoning whereas intellect seems to relate to a capacity type notion or depth of knowledge.

In 2004, MIT's Media Laboratory initiated the "10x" effort with the goal of magnifying human abilities by an order of magnitude ("10 ×"), or more, along various cognitive and physical dimensions [14]. This paper also starts with the citation and perspective of Licklider's 1960 paper. This paper describes possible new symbiotic interfaces along several different themes:

- Perceptual computing--signal processing and pattern analysis techniques for sensing and interpreting the environment, with a particular emphasis placed on interpreting the presence, identity, and activities of people
- Natural embodiment — human-friendly mechanical systems—basically an ability of computers to act on the world of interest to the human
- Natural representation—new frameworks for computational intelligence in which emotional states are connected to the physical—i.e. affect and embodiment influencing cognition
- Learning and expression — beyond programming--new ways of modelling semantics and social interaction for improved human-computer communication

That paper is an example of at least two issues involved in deciding an "appropriate" metaphor for symbiosis: (1) that there are (of course) fundamental differences between the computer as machine and the human as a biological entity and that addressing these is one vector of research, and (2) that both the human and the computer are ever changing in various ways that affect the nature of and possibilities for symbiosis. Humans today are "digital natives" that have overcome certain aspects of the human-computer impedance mismatch, and as well computers have become different types of devices, able to talk, to understand written input, etc. So we argue that symbiosis has contextual/temporal factors that influence the framing of the nature and potential for these symbiotic relationships.

Another example of this changing and always-dynamic relationship is in thinking about what has come to be called human computation [15], one definition of which is ""…a paradigm for utilizing human processing power to solve problems that computers cannot yet solve" [16]. This notion describes a particular type of symbiosis that can involve crowdsourcing, notions of collective intelligence etc as depicted in Figure 1 [15]:

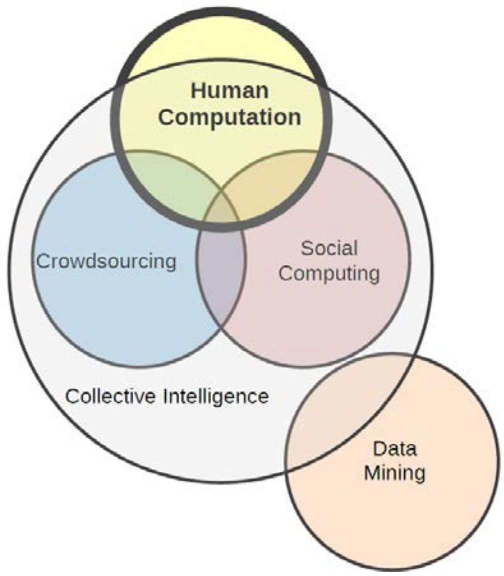


Figure 1 Human computation is a means of solving computational problems

Quinn and Benderson offer an extensive taxonomy of both this domain. We mention this because it shows how the changing nature of computation, as today being fundamentally distributed and shareable, offers new ways of solving problems, and new challenges and opportunities for human-machine interaction and synergy.

A key issue is how to develop a design approach to a human-machine dynamic that explicitly addresses design factors that influence symbiosis and moreover can improve symbiosis. Our early literature review reveals some controversy about such design approaches; the human factors engineering community has evolved the "user centered design" and "human centered design" paradigms that have an extensive literature [17] but these paradigms remain the subjects of considerable discussion. Norman, the author of what many consider to be the bible of human centered design in fact authors an article titled "Human-centered design considered harmful"[18] . Important in all this is a sense of perspective; some approach the formulation of approaches to achieving symbiosis from an augmented cognition point of view, a neuroscientific point of view, a telepathic point of view, etc. For this study, these are considered far too ambitious. We prefer the idea of "activity centered design". As our effort focused primarily on the central themes of analysis paradigms and technologies, we did not fully address the design of the human-machine interface and the addressing of the symbiosis issue. However, as will be seen, our primary thrust along this line was fully directed to reducing the human cognitive workload, which we see as a major impediment to the realization of an effective yet efficient basis for computational support to modern analysis. As will be pointed out in the sequel, sensibly all current capabilities to include the most modern still demand considerable human cognitive effort for early, front-end type processing; our approach proposes exploitation of named technologies for the construction of much more automated front-end processing, and is a unique aspect of our proposed design.

4. Goals and Requirements

In this study-based program, we sought to explore a number of possible computationally-aided enhancements in the ways that technologies can better support and improve the rigor and efficiency of intelligence analysis through the integration of new computationally-based methods and algorithms but also by exploring and nominating new ways in which improved human-machine symbiosis can be realized. Also, we were trying to strike the best balance between technologies and methods that are of the basic research variety while having plausibility in terms of potential for mid-term type operational deployment. Another main goal was toward providing support that can yield the type of "story" or narrative type product that many intelligence analysis environments require. These are those environments that allow for more contemplative methods, accommodating the formulation and evaluation of optional interpretations that have to be weighed and evaluated or argued for. This goal imputes a requirement for capabilities that support what we are calling "hypothesis synthesis" or "hypothesis fusion" as mentioned previously, where competing hypotheses that evolve either directly from evidence or developed from evidence or assumptions by individual, thematically-oriented support tools are traded off and synthesized into a defendable integrated hypothesis at the narrative or situational level. A major goal is to develop a design whose overall rationale is traceable to and consistent with joint service and Intelligence Community future directions in methodological development balancing effectiveness, efficiency, and rigor; as a result, we have made efforts to garner real-world viewpoints on these directions.

5. Future Directions in Intelligence Analysis

5.1 Reviews of Open Literature

The proposal for this grant was in fact partially inspired by our prior exploration of the nature of modern-day computational support for intelligence analysis in the open literature as summarized in [10]. That work extensively examined much of the literature on such techniques with a focus on technology strategies and interfacing strategies in regard to methods to achieve some level of symbiosis. It should be noted that this survey also collected works from the field of criminal analysis and the related area of Artificial Intelligence and the Law. Our research team at the Center for Multisource Information Fusion has also addressed these topics under a large Army Research Office grant for Unified Research on Network-based Hard and Soft Information Fusion (e.g., [19, 20]) for the Counterinsurgency domain. In both of these surveys, what we primarily saw was a strategy for analytical tool suite design that resulted in collages of disparate tools of various descriptions. Each of these tools can be argued to be individually helpful, producing what we called “situational fragments”, i.e., hypotheses, each of which are hypotheses about a particular slice of a situational condition. These problems, and the employment of modern technologies that allow evermore data and information to be available, are extraordinarily complex and it is natural to see ‘divide and conquer” solution, tool, and visualization strategies being applied. But the latent challenge for sensibly all human analysts involved in these situations is to connect the dots, evolve the most plausible story/narrative, or the most plausible argument in the face of inherent complexity and “big data” quantities and varieties of information. For that type of capability, we saw nothing at all in this survey, leading to our conclusion that there is a significant need for development of both a paradigm and associated technological support for hypothesis synthesis or fusion, aiding human analysts to assemble a more holistic picture (a narrative or story) much more efficiently.

5.2 Interactions with the National Air and Space Intelligence Center (NASIC)

In the Fall of 2015, coordinated by Lockheed staff, a team visit to the National Air and Space Intelligence Center was carried out. In particular, we visited the Advanced Analytics Cell (AAC) and were hosted by Mr. Hal Moon. We discussed the general approach of this effort and provided an overview briefing of our work to that moment. In turn, Mr. Moon provided an overview of AAC’s efforts toward developing improved analytical methodologies. There was considerable commonality in the respective lines of thought across the two activities, and the visit broadly provided a level of confidence that this project’s directions were sound and resonated with current advanced thinking at least in the Air Force as regards methods and needs of modern intelligence analysis.

5.3 Analytical Rigor in Intelligence Analysis/Argument Mapping

Another touchstone for the project as regards vetting our thinking and approach involved discussions with staff from the Army Intelligence Center at Ft. Huachuca, NM. Messrs Robert Sensenig and William Hedges (of Chenega Corp, advisors to the Army on intelligence matters) were our key points of contact. Two main topics were discussed: rigor in analysis, and the use of argument-based techniques of analysis. The Army is quite keen on the entire issue of improving rigor in analysis; this viewpoint certainly is consistent with our own thoughts regarding improvements in the intellectual aspects of analysis. Mr. Sensenig provided the charts of Fig s 2 and 3 below that depict the mapping/cross-correlation of analysis functions and levels of rigor, notionally showing an analyst’s mind-set across these functions and levels, as well as thumbnails of analysis activities across the matrices. These charts are among the resources we used to direct our efforts.

Low Rigor	Moderate Rigor	High Rigor
Hypothesis Exploration *"I have one hypothesis I like."* • No consideration of alternatives. • Argues how data that does not fit or is new can fit favorite hypoth.	*"I feel comfortable that one explanation accounts for majority data."* • Unbalanced focus on ML COA. • Acknowledges other COA possible. • Considers risks of alternative COAs.	*"I am confident of the best explanation and have seriously considered other possibilities."* • Interactive debate from multiple perspectives on alternatives. • Actively considers and tracks data that does not fit ML or MD.
Information Search *"I found something reasonably Comprehensive and believable."* • Did not go beyond routine sources • Did not select multiple sources. • Relied on second and third-hand sources, no direct comms with primary sources.	*"I am seeing repeating patterns, and they all seem to agree or there seems to be two primary possibilities."* • Actively seeks info that is not easily retrieved or collected. • Multiple data types and proximal sources considered for key findings • Read beyond specific tasking	*"I am not learning anything new. I reached theoretical saturation."* • Support from others to broaden sampled space. • Multiple data types and proximal sources considered for all inferences • More knowledgeable about subject area than most document authors.
Information Validation *"I found one that sounds good"* • Copies report with little re-interpretation, correlation • Does not display healthy skepticism. • No tracking of process, no knowledge of data pedigree	*"I verified my key arguments and predictions are based on the most trustworthy source I have"* • Attempts to verify arguments from multiple independent sources • Aware of how analysis could be wrong based on experience or feedback • Aware of corrupted data sources	*"I feel confident that I validated, by reasonable means, the facts used to support key arguments."* • Systematic, semi-formal processes employed to verify information • Clear distinction between facts, assumptions, inferences • Fully investigated "sourcing" 1

Figure 2 Mapping of Analysis Functions vs Levels of Rigor (Part 1)
(Courtesy of Mr. Robert Sensenig, Chenega Corp)

Low Rigor	Moderate Rigor	High Rigor
Inference Resilience *"My story/explanation/argument seems reasonable to me, independent of available supporting evidence."*	*"I feel that the evidence is reasonably solid for my primary explanation."* • Considers whether being wrong about some inferences would influence or negate the best explanation. • Beware false precision!!	*"I feel comfortable that the key Inferences are resilient to inaccurate Information."* • Uses strategy to systematically consider strength of evidence if individual interpretations debunked. • Actively looked for reasons why a source might misinterpret or manipulate data/information.
SME Collaboration *"I trust my supervisor to cover specialist content area or to be the SME."*	*"I have talked to SMEs, as time allowed, within my personal network."* • Attempts to consult some of the right people.	*" Leading expert in the key content area." (Beware Group Think!!)* • Capital expended to gain access to leading experts in multiple fields related to the analysis.
Information Synthesis *"I compiled the relevant info."* • Numerical values or graphs disconnected from key arguments.	*"I provided insight that goes beyond the source reporting & key documents.* • Validation of events in context. • Understanding depicted as an integrated view including tradeoff dimensions. (Frameworks, models).	*"I considered diverse interpretations trying to identify new concepts* • Sensemaking metrics are high. • Collaborative cross checks applied to data synthesis processes • Collaborative use of diagrams to show relationships between evidence and hypothesis. 2

Figure 3 Mapping of Analysis Functions vs Levels of Rigor (Part 2)
(Courtesy of Mr. Robert Sensenig, Chenega Corp)

Mr. Hedges recounted his experience in learning of argument-based methods of analysis and also shared segments of the Army's training activities in the teaching of argument mapping for intelligence analysts. Figure 4 shows an excerpt of one of the training segments directed to teaching of argument mapping.

U.S. ARMY INTELLIGENCE CENTER AND FORT HUACHUCA
Fort Huachuca, Arizona 85613-7002

LP Narrative & Teaching Plan: Argument Mapping
24 April 2013
PFN: xxxxxxxx

Enabling Learning SLIDE 2: Objective

ACTION:	Create an Argument Map to make analytic assumptions, intelligence gaps, or arguments more transparent.
CONDITIONS:	Given all class handouts to date, appropriate references, an operational framework scenario, and in-class discussion.
STANDARDS:	Create an argument map that incorporates critical and creative thinking and basic and diagnostic structured analytic techniques in order to provide clearer ACH understanding and validate the ACH.

Figure 4 Sample of Curriculum at Army Intelligence School Training in Argument Mapping (Courtesy of Mr. William Hedges of Chenega Corp.)

Overall, we believe it is quite clear that the thinking and approaches of this program are very consistent with modern thoughts in both the Air Force and Army in regard to:

- The use of improved intellectual strategies and methods
- The need for an movements to improve analytical rigor
- The employment of argumentation-based methods and technologies as one framework to achieve these goals

6. Approaches to Computational Support

6.1 Paradigms and Methods

In today's open-world environment, historical paradigms and methods that rely on deep analysis of an adversary's Tactics, Techniques, and Procedures (TTP's) as a basis for paradigms that can be broadly labeled as of a template-matching type are considered unworkable. Modern-day adversaries and problem conditions demand more flexibility and accommodation of imperfections in analysis techniques. These environments, that we call "weak knowledge" problems, require a more flexible approach and one that allows for unknown states of affairs and degrees of ignorance while carrying out the best analysis possible. Such methods are usually labeled as defeasible and abductive[1] and are directed to the most rational hypotheses

[1] We like Stanford's definition here (http://plato.stanford.edu/entries/reasoning-defeasible/): "Reasoning is *defeasible* when the corresponding argument is rationally compelling but not deductively valid. The truth of the premises of a good defeasible argument provides support for the conclusion, even though it is possible for the

that can be defended in some way as "best". In our exploration of alternatives, we narrowed our choices based on two factors: one was the commentaries on intelligence analysis and associated assertions about methodological requirements that balance evidence, arguments, and stories (i.e., nominated hypotheticals), and the other was a body of work we discovered that was centered in Europe that focused on methods of this type, with a deep basis on argumentation-based principles. One clear example of these remarks is shown in the writing of Schum [21] who suggests that:

- "**Careful construction of arguments** in defense of the credibility and relevance of evidence **goes hand-in-hand with the construction of defensible and persuasive narratives.**"
- "In constructing a narrative account of a situation of interest we must be able to **anchor our story appropriately on the evidence** we have that is relevant to the conclusion we have reached. Careful argument construction provides the necessary anchors."

These remarks, and the results of our surveys, suggest an exploration of methods that jointly exploit the union of evidence, arguments, and stories, in a synergistic dynamic that leads to "best" narratives that holistically convey the most rational explanation of the evidences and sub-stories. These source materials were the foundation of the evolution of our thinking to explore a paradigm of this nature.

6.2. Argumentation Methods

As we contend above, one main technological/theoretical theme that we pursue here is the examination of argumentation-based concepts, methods, and computationally-supported tools as one candidate paradigm supportive of intelligence analysis. Argumentation-based methods have a long history in the law and in the teaching of critical thinking, and in the last decade or so have found their way into supporting criminal and intelligence analysis. These extended applications have largely been a result of research and development in the construction of computational tools for "diagramming" or "mapping" arguments that enable and streamline the examination of the veracity of pro and contra arguments in various situations[2]. Before reviewing the state of the art in computational methods for argumentation based reasoning, we briefly review the different paradigms for argumentation itself; that is, there are different flavors or variations of methods that have the core notion of an argument as their foundation. This summary review is shown in Table 1:

premises to be true and the conclusion false. In other words, the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information."

[2] By the way, we see the (necessary) balancing of Pro and Contra arguments as another good feature of these argumentation methods; to some degree this is a built-in preventative to the human foible of confirmation bias.

Argumentation types	Methods	Prototypes[3]
Abstract Argumentation	**formal logic, theorem proof, and based on the notion of argument acceptance and attack)**	
Epistemic Practical Assumption-based	Deductive reasoning to support beliefs Abductive reasoning to support actions Arguments are deductions based on a set of assumptions and inference rules	CISpaces, Carneades Araucaria and Various others
Hybrid Methods	**Combination of logic and probability or belief**	
Assumption based probability/belief based argumentation. A probabilistic extension of abstract argumentation.	Conjunction of uncertain assumptions to define arguments and disjunction of arguments Assigning probabilities to arguments and defeats	ABEL
Belief-story Based	Observations are explained by hypothetical stories. Uncertain arguments based on evidence are combined to support alternative stories and select the most probable one (abductions)	This NPS Project's Goal System, Under development

Table 1 Types of Argumentation-based Paradigms

Largely, the focus of much of the research on argumentation schemes is based on arguments that have a formal, first-order logic basis (Abstract Argumentation in Table 1); this version of argument based reasoning allows for the employment of all the rigor of first-order methods but as we have claimed, our focus needs a more flexible approach. Relaxations of the first-order requirements are achieved by other methods such as assumption-based and probabilistically-based variations. These however, at least in the literature reviewed, still mostly focus on closed-world problem domains. We seek methods that combine certain desirable capabilities:

- Allowance for open-world reasoning
- Allowance for assigning of and related computations about Beliefs in arguments (i.e., a basis for assigning and combining/propagating uncertainty)

[3] See later discussion on Prototypes for citations.

- Integration of human intelligence that enables hypothetical stories to be combined with hypotheses resulting from evidence-based arguments

Such methods need to be labeled “hybrid” (as in Table 1) since there are no formalisms in the argumentation taxonomy that would assign a specific name to such methods. Abductive reasoning is often labeled as “backward reasoning” in that it explores/nominates plausible conclusions or assertions that can “explain” or rationalize the evidence available; the notion is that a rearward look is taken from the conclusion toward the available evidence. Abductive reasoning is also often described as reasoning to the best explanation. Our approach is also hybrid in bringing together the abductive reasoning over both the uncertain arguments and human-nominated storylines and rationalizing both lines with the also-uncertain evidence. To deal with these uncertainties, we propose to incorporate the Transferable Belief Model (TBM) proposed by Smets [22]. TBM is elaborated on in Section x of this report but briefly the TBM is a multi-parameter model, in which quantified beliefs in hypotheses about an object or state of the environment are represented and combined at the *credal* level[4] while decisions are made based on probabilities obtained from the combined belief by the *pignistic*[5] transformation at the *pignistic level*. So taken together, the proposed approach can be summarized as involving the explicit incorporation of uncertainty into hybrid story-based argumentation, depicted in Figure 5:

[4] Credal will be seen to mean belief but in regard to conducting analysis this term is taken to mean a (human’s) conviction of the truth of some statement or the reality of some being or phenomenon especially when based on examination of evidence.
[5] Pignistic is a term coined by Smets and is drawn from the Latin pignus for “bet”, and can be taken to imply or relate to a probability that a rational person would assign to an option when required to make a decision.

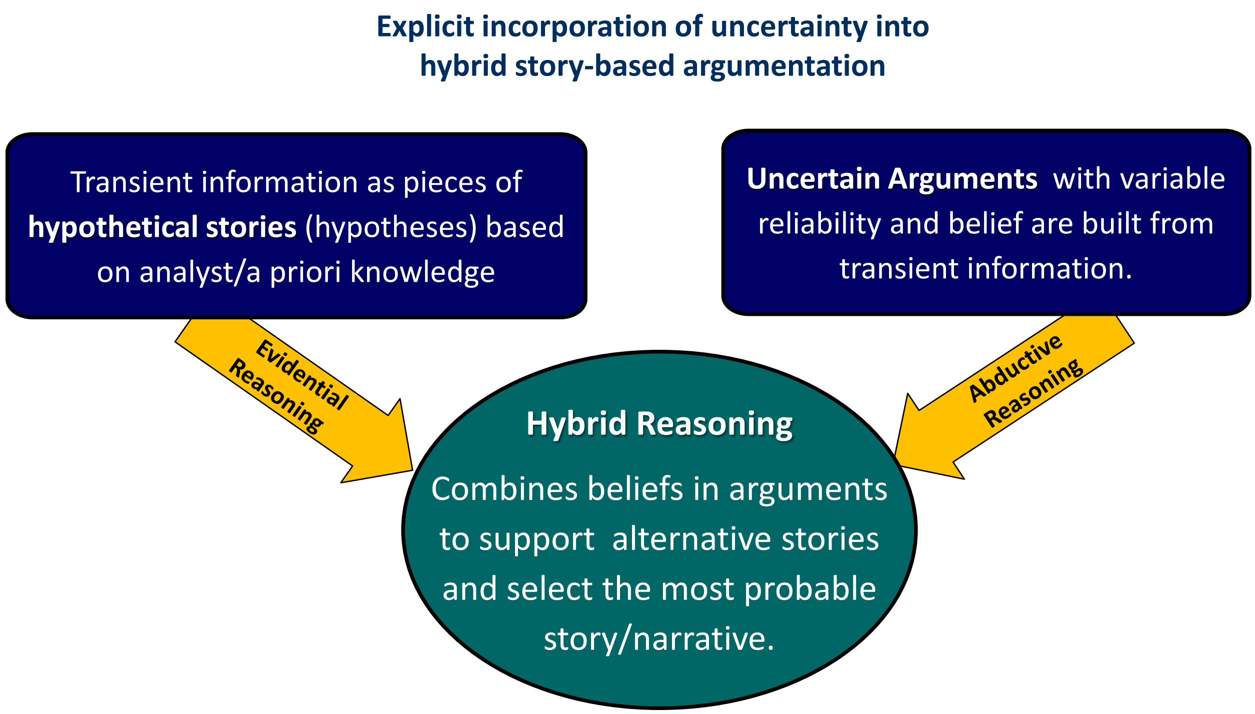


Figure 5 Depiction of the Proposed Hybrid Approach

It is important to note that there has been development of some of the formalisms associated with hybrid argument-story-based reasoning that is one important basis of our approach. Floris Bex, a professor in the Department of Information and Computing Science of the University of Utrecht, the Netherlands, has written several papers describing the bases of these ideas on combining and exploiting these two lines of reasoning (eg [23-25]). The basic ideas are shown in Fig 6 that shows that:

- Arguments are derived from evidential foundations
- Stories are analyst-nominated (with computational support, eg prior case libraries) hypotheticals
- Together these lead to the assembly of sub-stories and, again with computational support (see Section x on our ideas), to the development of an integrated Narrative/Story

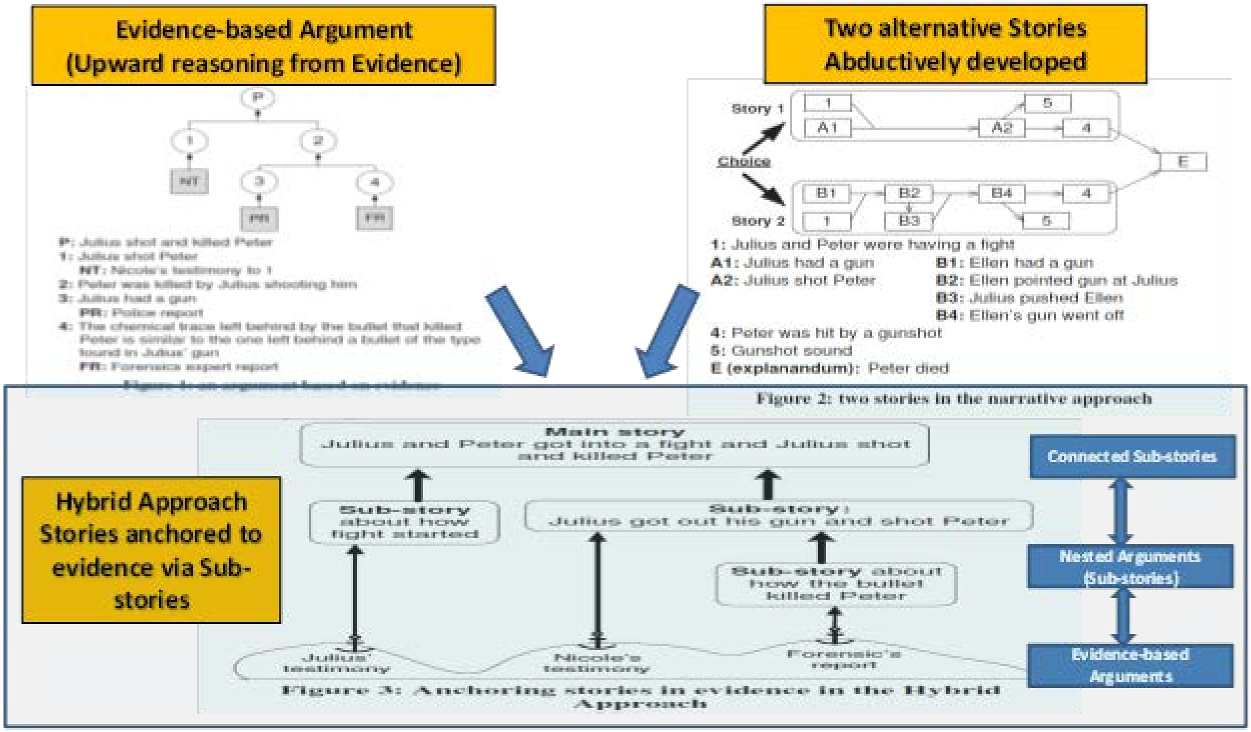


Figure 6 Overview of Bex's Scheme for Joint Argument-Story Exploitation

In the following, we provide our view of the state of the art in each of several functional areas necessary toward realization of a desired level of automated capability for a future semi-automated, computationally supported analysis prototype that realizes the hybrid capability described. We note, from the literature, a set of particular argumentation-related functional categories: Argument Detection-Construction-Invention-Mining-Accrual and, importantly (as it dominates the literature) Visualization that will serve as the basis for our review.

6.3 Computational Support to Argumentation: The State of the Art

It is realized that the input to any modern intelligence analysis system could be in a wide variety of formats and types in terms of media and modalities. As regards the role of these varying inputs toward supporting argument formation however, it is considered that textual inputs provide the most likely format for somewhat-direct input-to-argument formulation. Most other input types would more likely represent evidential data (such as sensor data) and require a more complex structuring process to frame the data into argument forms. (Later it will be seen that address sensor data as an input stream of interest in proposing our design, but it will be seen that sensibly all current systems do not include such "hard" data as input.)

As will be seen in our review of current prototype argument systems, the front-ends of these prototypes do not currently provide any automated support to the identification of either the basic linguistic form of an argument (based on lexical content and other factors) nor types of argument structures based on argument taxonomies (usually called "schemes" in the argument literature) from textual report, prose-type input, whether structured or not. Thus, a significant human cognitive operation is needed in these prototypes for the formulation of the very basic constructs (arguments) upon which next analysis steps, some computationally-aided, depend. Moen et al [26] in discussing the Araucaria prototype designed for argument visualization, say that "The manual structuring of an argumentative text into a graph visualization as is done in the Araucaria research is a very costly job. "

However, we will see that computational support for extracting parts of or entire argument schemes from text has been addressed but has not, for whatever reasons, been integrated into modern prototype systems. As noted above, this functional activity comes under different names, such as argument detection, argument construction, and argument mining—we simply use the term detection here but draw on works having these other labels to describe what is happening in the research community. We will review some sample works in this area and also provide a broader summary view of the state of the art.

For the Reader: our reviews are running commentaries about selected papers from the literature that address each reviewed topic; in various places any emphasis provided is our own. Some excerptions from the original papers are included without quotation as we see this report as a project technical report, not a public document.

6.3.1 Argument Detection

- **Moen, M., et al, Automatic Detection of Arguments in Legal Texts, [26]**

This paper describes the results of experiments on the detection of arguments in texts with a focus on legal texts. As will be seen in related works on detection, the detection operation is seen as a **classification problem** based on defined features of a postulated argument scheme. A classifier is developed in the paper and trained on a set of annotated arguments. Different feature sets are evaluated involving lexical, syntactic, and semantic and discourse properties of the texts, and each of their contributions to classifier accuracy is examined.

Strategies for detecting argument constructs clearly require some defining process for the nature of argument forms or schemes in a linguistic sense; said otherwise, an ontology of argument forms is required. Citing [2], Moen states that "The most prominent indicators of rhetorical structure are lexical cues [27], most typically expressed by conjunctions and by certain kinds of adverbial groups." Humans can do this well but one important factor exploited by humans to do so is the context of the textual phrases, and this is very hard to do automatically. The approach in [26] is admitted to be a bounded first step toward automating this process, and they take an approach built on isolated sentences. They represent sentences as a vector of features and use **annotated training data** to train a classifier. (It will be seen that this problem is broadly treated as a classification problem in the literature.) We will not review the details of the features and methods but they use a **multinomial Bayes classifier** and a **Maximum Entropy based classifier** in this work. It is interesting to see that even **simple feature sets yield reasonable (~70+% accuracy) results.** The paper also reviews related works and remarks that this type of research on detection is very limited in the legal domain at least (as of the date of [26], 2007).

.

- **Mochales-Palau and Moens [28]**

In a later work, Mochales-Palau and Moens [28] develop an approach to detect sentences that contain argument structures (apart from, ie not discerning the existence of Walton-type schema; see below on this issue and [294] regarding the schema). A **maximum-entropy-based classification approach** is used to determine if input sentences are argumentative or not, and more specifically if they contain a premise, a conclusion or a non-argumentative sentence. These same authors also study and develop a context-free grammar for argument detection in [30], but this was a very limited study across a 10 document training set.

- **Feng and Hirst, Classifying Arguments by Scheme, [31]**

This work is oriented to a subtle issue in argumentation, the issue of **enthymemes**; as part of an approach to argument detection, in reasonably-frequent cases, there are **implicit premises** that are never present in the prose text, and these are called enthymemes. To do this however, they argue that by first identifying the particular argumentation scheme that an argument is using will help to bridge the gap between stated and unstated propositions in the argument, because each argumentation scheme is a relatively fixed “template” for arguing. The idea here is that the argument scheme classification system is a stage following argument detection and proposition classification; that is, a two-stage system involving two different classification systems.

This paper (and some others) relies on the notion of **argument schemes or schemata**; such schemes are structures or templates for forms of arguments. Most argumentation schemes are for defeasible arguments. **Walton’s set of 65 argumentation schemes** [29] is one of the most-cited scheme-sets in the argumentation literature. According to [31], the five schemes defined in Table 2 copied below are the most commonly used ones, and they are the focus of the scheme classification system that is described in this paper. The functional approach is shown in Figure 7, where it can be seen that argument detection from text precedes the argument *scheme* classification step.

Argument from example *Premise*: In this particular case, the individual a has property F and also property G. *Conclusion:* Therefore, generally, if x has property F, then it also has property G.
Argument from cause to effect *Major premise*: Generally, if A occurs, then B will (might) occur. *Minor premise*: In this case, A occurs (might occur). *Conclusion*: Therefore, in this case, B will (might) occur.
Practical reasoning *Major premise*: I have a goal G. *Minor premise*: Carrying out action A is a means to realize G. *Conclusion*: Therefore, I ought (practically speaking) to carry out this action A.
Argument from consequences *Premise*: If A is (is not) brought about, good (bad) consequences will (will not) plausibly occur. *Conclusion*: Therefore, A should (should not) be brought about.
Argument from verbal classification *Individual premise*: a has a particular property F. *Classification premise*: For all x, if x has property F, then x can be classified as having property G. *Conclusion*: Therefore, a has property G.

Figure

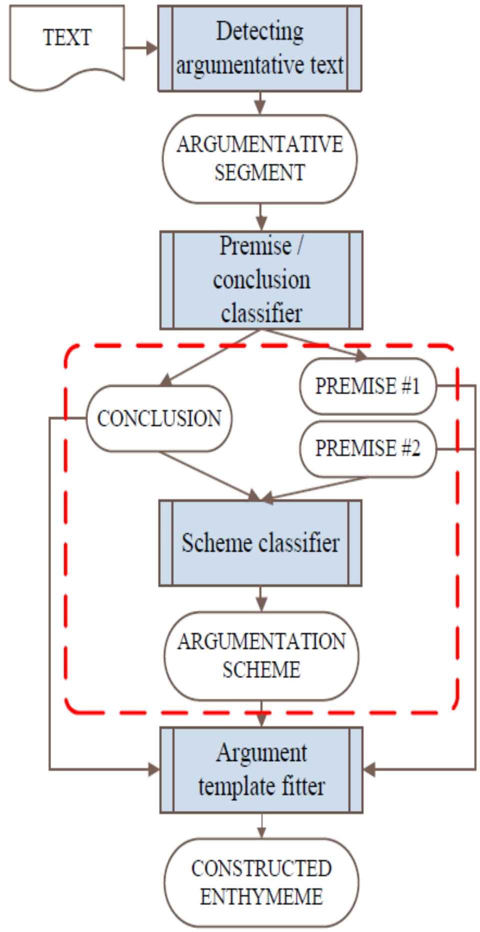


Figure 7 Functional Flow of Argument Scheme Detection [31]

Table 2 Five top argument schemata from [29], according to [31]

The classifier approach employs the C4.5 algorithm that is essentially entropy based. Performance is quite variable, since the various argument schemata vary significantly in the specificity of cue phrases; this is an issue to be dealt with in classifying argument schemata. Note that a training data set for either argument detection or scheme detection requires that the textual corpus be labeled with the “true” argument constructs. This study used the Araucaria data set available at the Araucaria research project website, http://www.arg-tech.org/index.php/projects/.

6.3.2 Argument Mining

Moens: State of the Art in Argument Mining [32]

Argumentation mining is defined in [32] as the (automated/automatic) detection of the argumentative discourse structure in text or speech and the recognition or functional classification of the components of the argumentation. It is clear from this definition that various functional capabilities are required in mining, to include detection of lexical units, identification of sentences containing arguments, and the fit of an argument sample to a predefined argument schema. This type of functionality can be said to fall into the domain of Information Retrieval systems, to provide the end user with instructive visualizations and summaries of an argumentative structure. Moens [32] dates argument mining as having started in 2007. The notion of argument "zoning" is mentioned as an area of some study, where a document or corpus is examined to localize sections possibly containing argument-based content. Moens reviews some works that perform these types of functions as typical of the current state of the art; typical Precision/Recall/F measures are in the high 60 to low/mid 70% range, which is just fair performance.

This paper also describes some capability goals for argument mining systems. While discussing the use of machine learning methods, the goal of detecting or recognizing a "full argumentation tree" is mentioned. Cited papers include [33, 34] that use either a set of piecewise classifiers or a single set-wise or tree-wise classifier, but these are cited only as methodological examples, ie these works do not apply such methods to the argument mining problem. Another important argumentation mining issue stated in [32] is the correct identification of the relationships between text segments (e.g., the relationship of being a premise for a certain conclusion) and defining appropriate features that indicate this relationship. Moens suggests that textual entailment in natural language processing, which focuses on detecting directional relations between text fragments may be useful.

6.3.3 Argument Invention

- **Walton & Gordon—The Carneades Model of Argument Invention [35]**

This paper seems a bit off-topic for our purposes but one aspect that may be of interest is that the mechanics involved in argument invention may hint at how Stories (in a knowledge base) and arguments an achieve some symbiosis. Argument invention is a method used by ancient Greek philosophers and rhetoricians that can be used to help an arguer find arguments that could be used to prove a claim he needs to defend. The Carneades Argumentation System (named after the Greek skeptical philosopher Carneades) is said in [35] to be the first argument mapping tool with an integrated inference engine for constructing arguments from knowledge-bases, designed to support argument invention. It can be said that the notion of invention revolves around the notion of how arguments are evaluated or defended; the idea is to provide automated support to improve the acceptability of an argument. This tool is intended for rhetorical-type applications but conceptually could have applicability in analysis frameworks.

We offer an aside re argument evaluation, drawn from [36], as follows: one approach to argument evaluation revolves around the idea of “critical questions” to evaluate an argument. In [36] we have: “Critical questions were first introduced by Arthur Hastings [37] as part of his analysis of presumptive argumentation schemes. The critical questions attached to an argumentation scheme enumerate ways of challenging arguments created using the scheme. The current method of evaluating an argument that fits a scheme, like that for argument from expert opinion, is by a shifting of the burden of proof from one side to the other in a dialog. When the respondent asks one of the critical questions matching the scheme, the burden of proof shifts back to the proponent’s side, defeating or undercutting the argument until the critical question has been answered successfully. At least this has been the general approach of argumentation theory.” Thus, the presence of critical questions could serve as a mechanism to assure that pro and contra sides of an argument receive attention.

The Carneades design approach provides a number of “assistants” for helping users with various argumentation tasks, including a “find arguments” assistant for inventing arguments from argumentation schemes and facts in a knowledge base, an “instantiate scheme” assistant for constructing or reconstructing arguments by using argumentation schemes, and a “find positions” assistant for helping users to find minimal, consistent sets of statements which would make a goal statement acceptable. The schemes representing knowledge of the domain in the knowledge base must be programmed manually by an expert. A distinctive contribution of the Carneades system is the integration of an inference engine in an argument mapping tool. Although the paper does not emphasize application in the legal domain, it seems clear that this system is oriented to either legal applications or in rhetorical applications as mentioned previously.

6.3.4 Argument Visualization (a.k.a. Mapping, Diagramming)

Argument visualization is often claimed to be a powerful method to analyze and evaluate arguments by providing a capability to understand dependencies among argument components of evidential components, premises, and conclusions, focusing on the logical, evidential or inferential relationships among propositions. Argument visualization and theoretical modeling play important roles to cope with working memory limitations for problem solving, providing some relief to the cognitive workload that these analyses impute. Since the task of constructing such visualizations (also described in the literature as argument mapping or diagramming) is laborious, researchers have turned to the development of software tools that support the construction and visualization of arguments in various representation formats that have included graphs and matrices among other forms. To say that there have been a number of prototype systems developed that support argument diagramming is rather an understatement—a website provided by Carnegie-Mellon University (http://www.phil.cmu.edu/projects/argument_mapping/) shows, just on the first page, the following subset of tools shown in Table X; the complete table goes on for 2-1/2 pages. Note also the range of representational forms, in part dependent on the argument-model used in the application.

The effectiveness of such diagramming or mapping tools is reviewed in [38]. Among the tools that were experimentally tested for their effectiveness were Belvedere, Convince Me, Questmap, and Reason!Able, which are a sampling of tools from Table 3. While there are many issues regarding such evaluations discussed in [38] to include criticisms about statistical testing methodology, the paper concludes that “most results indicated that the tools have a positive effect on argumentation skills and make the users better reasoners. However, most experiments did not yield (statistically) significant effects.” Another study [39] showed that (manual) argument mapping generally helped in understanding arguments and also enhanced critical thinking; the study also showed that the benefits were greater with computer based argument mapping.

In [40], Mani and Klein review structured argumentation as an analysis framework for “open-ended” (ie in operational cases where absolute truth is unknown) intelligence analysis. The paper is short, opinion-type paper and asserts that structured arguments are a means not just of representing and reusing reasoning (one useful benefit), but also a means of communicating and sharing the argument, as analysis is often collaborative. They suggest that one way of assessing the quality of the associated reasoning is to determine how easy the argument is to follow and understand. If arguments are constructed in agreeable ways (eg based on argument models/schema) and correspondingly visualized, presumably they can be more easily communicated with and discussed with others.

Table 3 Sampling of Computer-supported Argument Diagramming Tools
(see http://www.phil.cmu.edu/projects/argument_mapping/)

Tool	Description	Representation	Audience
Athena	Argument mapper from Blekinge Institue of Technology and CERTEC, Sweden.	Simplified Toulmin	Education
ArgMAP	Argument mapper	Simplified Toulmin	Research
ArguMed	Argument mapper based on DEFLog	DEFLog (Toulmin extension)	Research
Argutect	Argument mapping-like "thought-processor" from Knosis, Pittburgh.	Thought tree (tree of questions and answers, can be used as simplified Toulmin)	Productivity, Education
Araucaria	Argument mapper from Univeristy of Dundee, UK.	simplified Toulmin	Education
Belvedere	Collaborative concept mapper and evidence matrix originally developed by D. Suthers at LRDC, Pittsburgh, now at LILT, University of Hawai'i at Manoa.	Inquiry / Evidence Maps and Matrices (links between claims and supporting data)	Education
Causality Lab	Allows students to solve social science problems by building hypotheses, collecting data and making causal inferences.	Causal diagram and data charts	Education
Carneades (.pdf)	Toulmin based mathematical model for legal argumentation	Toulmin	Law
ClaimMaker/ ClaimFinder/ ClaimMapper	Concept mapping of knowledge claims from S. Buckingham Shum's Scholarly Ontologies Project, KMI, Open University, UK.	Concept map with semiformal ontology for argumentation	Research
Compendium	IBIS mapping tool orginially developed by Verizon Reserach Labs and associated with CogNexus Institue and KMI, Open University.	Dialogue map (concept map with ontology: nodes can represent issues, ideas, pro, con, and notes)	Ill-structured problems
Convince Me	Creates diagramatic representations of hypothesis and evidence	Evidence map	Education
Debatabase	"Debatabase is the world's most useful resource for student debaters. Inside you will find arguments for and against hundreds of debating Topics, written by expert debaters, judges and coaches."	Communal, simplified Toulmin	Education

To allow an appreciation for what such visualizations look like, we show some examples in Fig. 8; these are drawn from Gordon's presentation in [41]; we use his format as it typically provides a screenshot with some remarks on associated features. Belvedere and Araucaria are very frequently cited as exemplars of relatively recent prototypes for argument visualization (see for example [42, 43]). A yet more recent example is Rationale, developed in Australia [44], and using a new “hi-tree” approach to visualization. The most recent prototype we are aware of is CISpaces, developed under joint US-UK efforts and led by Norman at the University of Dundee.

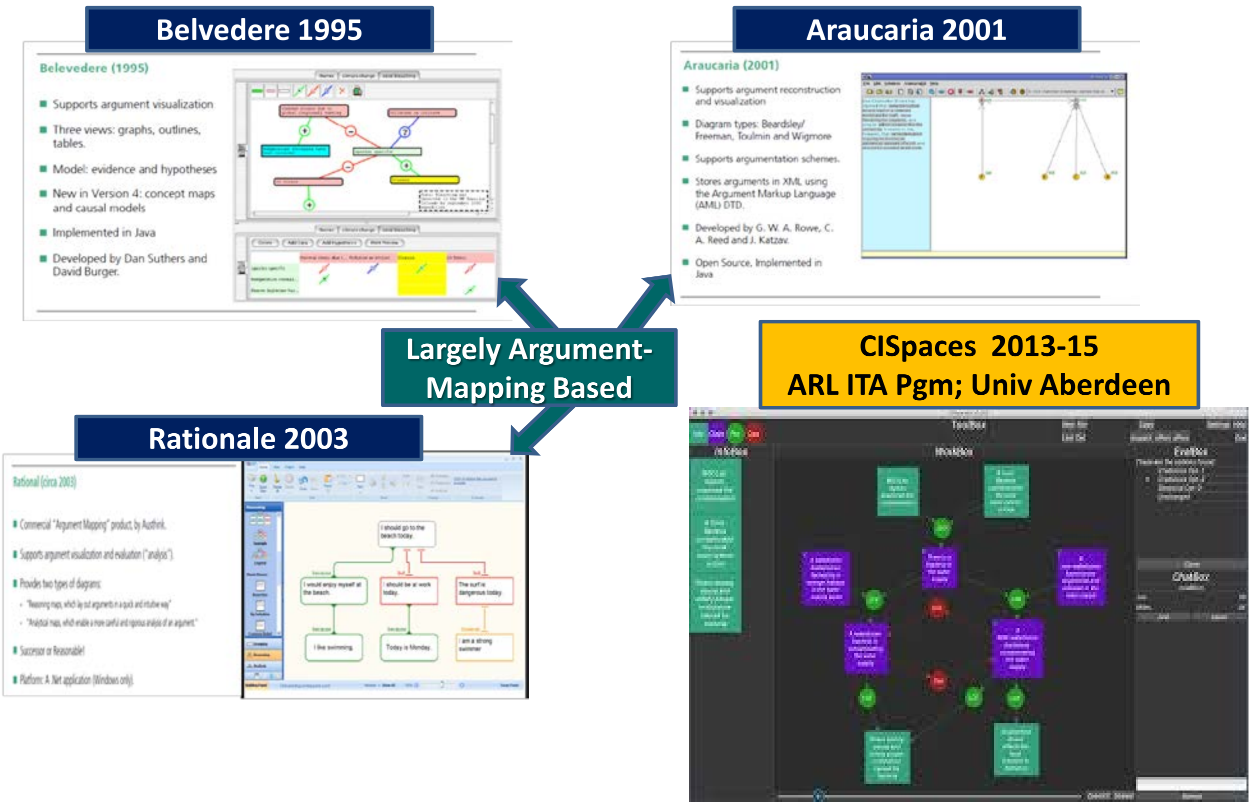


Figure 8 Sampling of Argument Visualization Prototypes

7. Current-day Computational Support to Argumentation

One other remark that we will offer here is that the greater proportion of research along the lines of computational support schemes for analysis has been carried out in Europe or at least outside of the USA. Among the leading centers of such research are:

- ARG-Tech, at the University of Dundee in Scotland (http://www.arg.dundee.ac.uk/)
- Centre for Research in Reasoning, Argumentation and Rhetoric, University of Windsor, Canada (http://www1.uwindsor.ca/crrar/)
- Intelligent Systems Group, University of Utrecht, Holland (http://www.cs.uu.nl/groups/IS/)
- Intelligent Systems Group, University College London(http://is.cs.ucl.ac.uk/introduction/)

Another barometer of this situation can be seen by examining the locations of the COMMA Conferences (Computational Models of Argument):

- 2006: University of Liverpool, (UK)
- 2008: Universite Toulouse, (France)
- 2010: Desenzano del Garda (Italy)
- 2012: Vienna University of Technology (Austria)
- 2014: Atholl Palace Hotel, Pitlochry (Scotland)
- 2016: Potsdam University, Germany

To the extent that there is belief that computationally-supported argumentation methods can be helpful to intelligence analysis, this situation should be of concern to the US academic and industrial research communities.

7.1 AVERS and CISpaces as Leading Relevant Prototypes

This research program was largely initiated by an early review of a dissertation in Holland having to do with "Sensemaking software for crime analysis" [45] by Susan van den Braack. That dissertation provided the spark of thinking, as was first explored in that work, for a hybrid, story and argumentation based approach to intelligence analysis since intelligence and criminal analysis requirements have quite similar requirements. This dissertation described AVERS as a prototype developed within the dissertation effort that was designed to explore alternative "scenarios" (stories in effect) based on evidentially-supported arguments. A prototype was developed in the university framework but unfortunately the code for that prototype was not subsequently maintained (we had contacted Dr. van den Braak to explore this). Nevertheless, as described in [46], it is clear that the thinking related to the design and realization of AVERS was very synergistic to our line of research. Formalisms for combining stories and arguments in this hybrid environment were put forward in [47].

During our program, largely because of our close relations to researchers at the Army Research Laboratory, we learned that, under the "International Technology Alliance (ITA)" program (a US-UK cooperative research program) that a team at the University of Aberdeen (at ARG-Tech as noted above) was carrying out the development of a prototype called "CISpaces", with goals also similar to ours.

CISpaces was conceptualized as an initial set of tools for collaborative analysis of arguments and debate, providing a uniform way of constructing and exchanging arguments based upon argumentation schemes. The top-level functional design is shown in Fig 9 below [48] and comprises three main services in a service-based architecture:

- the evidential reasoning service, supporting collaboration between users in drawing inferences and forming opinions structured by argumentation schemes;

- the crowd-sourcing service, enabling users to post requests for aggregated opinions from samples of a population;
- the provenance reasoning service, facilitating the storage and retrieval of provenance data including provenance of information and analysis.

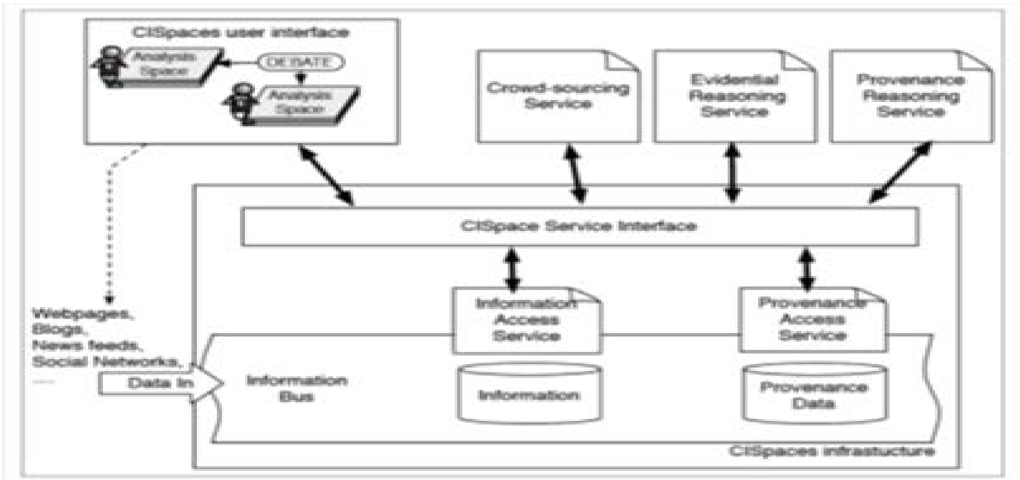


Figure 9 CISpaces Functional Architecture [48]

The core components of CISpaces, as it is highly oriented to a collaborative, multi-analysts environment, are the WorkBox, the ChatBox and the ReqBox. As described in [48], the WorkBox permits users to elaborate information by adding new claims or by manually importing information and conclusions from different locations; e.g., social networks, blogs. Different forms of argumentation-based dialogue are supported through the ChatBox: collaborative debate, information retrieval through crowd-sourcing, and reasoning about provenance. The list of active debates is intended to be maintained in the ReqBox. A snapshot of the analyst interface that shows these workboxes/services is shown in Figure 10:

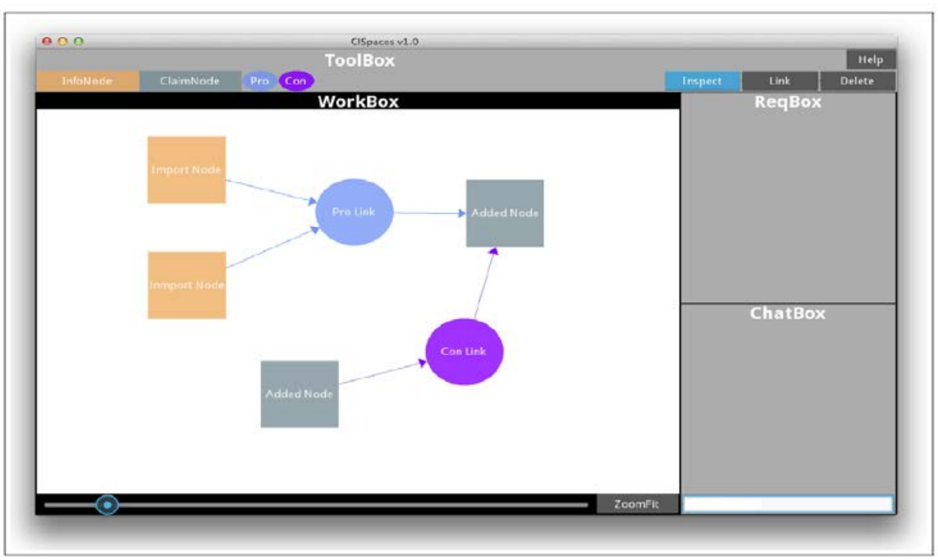


Figure 10 CISpaces Analyst Interface [48]

While the development of a real software prototype of this type should be applauded for its forward-thinking approach and for moving the bar of computational support to argumentation to a new level, our thoughts on prototype design addressed other, additional issues:

- Inclusion of both Hard/sensor data as well as Soft/textual/linguistic data as input
    - This is a major change as sensibly all existing argumentation support prototypes are strictly text-input-based
- Major reduction in analyst cognitive workload
    - We see this as involving an aggressive inclusion of front-end, automated processing to aid in argument detection and construction, a major cognitive workload factor of all current prototypes, to include CISpaces.
    - Another aspect is in automated support to final analysis product development, seen as a narrative or story descriptive of a situational estimate of interest (none of the computational systems described here address this at all)
- Major concern for managing information quality along various lines, including automated support for relevance-checking and tracking and assessing provenance of input sources.

Because of our concern for these information quality factors, we established a research thrust along these lines, summarized in the next section. A later section also addresses our ideas, largely from our Lockheed teammates, on computational support to narrative development.

8. Foundational Issues in System Design: Focus on Quality in the Large

In this overall effort, our approach has been intended to be holistic in trying to conceptualize a total-system type capability and to evolve the associated functional design. We also have given thought to the "meta-qualities" of any such design and initiated an effort within the program to examine what we are calling "Foundational Issues". What we mean by this are the fundamental aspects of Information Quality, which include notions of Relevance and Uncertainty. This section addresses the results of our probe into these topics; it will be seen that such issues have been addressed in our functional design.

8.1 Information Quality Effects on Decision-making with Argumentation.

There are multiple definitions of information quality in the human-machine environment such as "Quality is the totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs" [49], "Quality is the degree, to which information is meeting user needs according to external, subjective user perceptions" [50], and "fitness for use"[51]. It can be seen from these definitions that quality is measured in terms of potential and actual benefits to the user (in the human-machine environment users can be humans or automatic processes). However the assessment of the "fitness for use" is based is on the characteristics of information representing inherent properties of information. The inherent information characteristics (information about information) represent *objective quality* or *meta-data* [52]. The same value of a quality characteristic can represent both meta-data when it is considered by itself but becomes subjective when considered in relations to use's objective in a specific context. For example, timeliness can be either a number between the actual and expected arrival time of the information or measure of usefulness of this information for the user's decision. Information quality has to be evaluated at each step of information exchange in the system to decide whether this information is useful to the interim processes and decision maker.

Figure 11 shows a subontology of information quality, which contains characteristics especially important for building a human-machine belief based argumentation system. The next sections will describe these characteristics and incorporation them into belief-based argumentation in more detail.

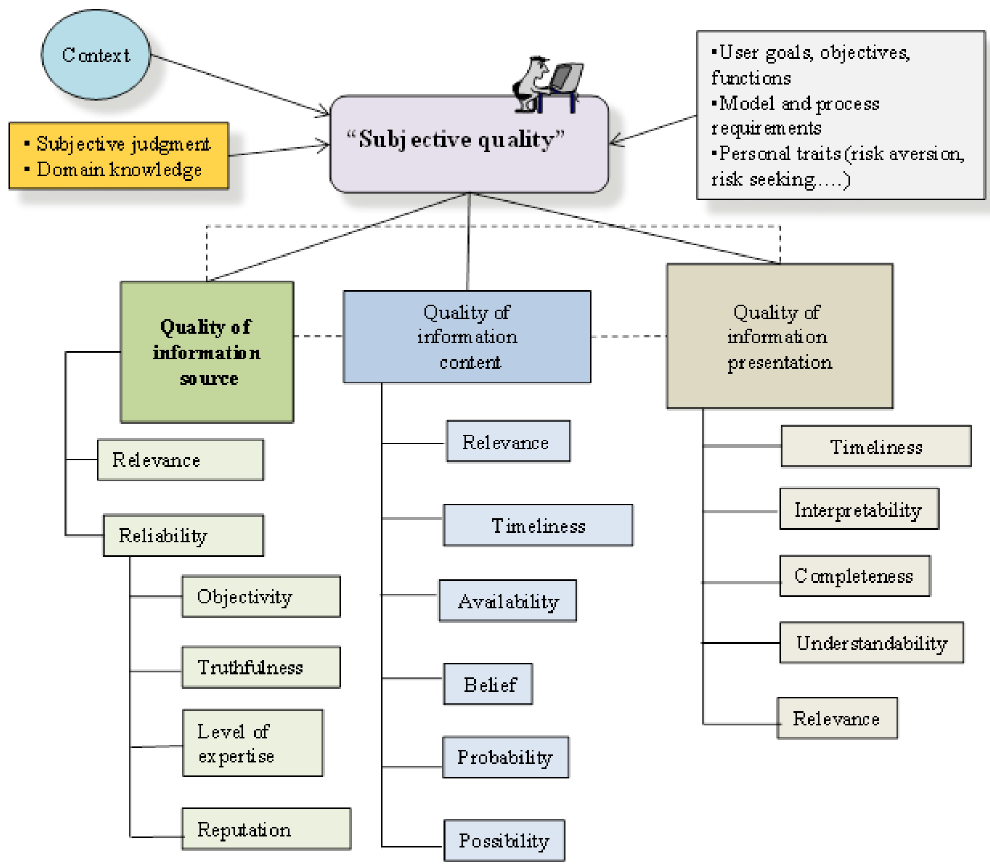


Figure 11. Ontology of the most important for argumentation subjective and objective quality characteristics (based on [52])

8.1.1 Relevance

As we can see from Figure 11, relevance is one of the central quality characteristics, which can characterize quality of information content, information source and information presentation. In order to incorporate models defining relevance in a mixed-initiative argumentation system it is necessary to understand what relevance and its properties are, the methods of determining the level of relevance, and how to measure its effect on performance of the system and its elements. The notion of relevance has been discussed in many publications in philosophy, law, pragmatics, and information retrieval but the problem of relevance in computational models of argumentation for threat assessment has not received much attention. At the same time the process of relevance evaluation at many steps of decision making by argumentation is similar to the processes of information retrieval.

According to the definition of the Meriam-Webster dictionary, information is relevant if it has "significant and demonstrable bearing on the matter at hand." Thus relevance is not a property but "is understood as a relation; relevance is a tuple—a notion consisting of a number of parts that have a relation based on some property or criteria" [53]. Formally a tuple is ($\{P_i\}$, R, $\{Q_j\}$, S,C), where $\{P_i\}$ and $\{Q_j\}$ are sets of both tangible or intangible objects, and R is a criterion defining relevance of these sets (e.g. utility), and S is a measure of the strength of relevance. If $S=0$, $\{P_i\}$ and $\{Q_j\}$ are not related, if $S=1$ they are completely related. In the uncertain environment relevance is not binary and $S \in [0,1]$. Relevance strongly depends on context as well as goals, functions, and expectations of decision makers. The dynamics of context, goals, and functions of the decision makers in the dynamic environment make relevance a temporal attribute: irrelevant information can become relevant later or relevant information can become obsolete at a certain time. There are several questions to be answered before a piece of information can enter the system:

- Is it relevant to the task or purpose of the processes? To what extent?
- Is the level of relevance enough to justify the use of this information?
- How reliable or trustworthy is the source of the information?
- Whether the information arrives in time?

Thus relevance of the information content depends on the reliability of the source of the information as well of its timeliness. Relevant information coming from a source of low reliability (broken sensor, or malicious human source) is irrelevant. Thus relevance of information content has to be evaluated along with other characteristics of information quality.

There is a definite connection between the cognitive effects of information, information processing time, and relevance [54]:

- The greater the cognitive effects, the greater the relevance is.
- The smaller the processing effort required for deriving these effects, the greater the relevance is.

Taking into account relevance of information has become more important with the increased role of social media as a source of information, which significantly increases the amount of information to be considered, which in turn increases cognitive overload of analysts. Incorporation of irrelevant data into fusion processes not only can increase cognitive overload and skew the quality of the result of this process but also negatively affects the performance of the other processes and ultimately decision making. In a human-machine system consideration

of relevance as a characteristic of the quality of information presentation is especially important to increase the information cognitive effect.
A relevance filter in an argumentation system should be considered for any elements of argumentation such as story, premises, conclusions, hypotheses. In a human-computer system arguments can be created by both human and automatics processes.

First relevance has to be considered for pieces of transient information before it enters the system in order to evaluate how relevant this information is to the goals, objectives and functions of the analyst even if this information does not contain arguments. At this point information is considered in relation to essential elements of information defined by these goals, objectives and functions. Information relevance or irrelevance here can be considered. Information relevance means that given the value of Z in context C, obtaining information about X gives us no new information about Y [60]. Since there may be multiple analyst/automatic process sub functions, relevance of this information need to be evaluated according to each of them. For example since threat is characterized as an integrated whole of threat, opportunity, and capability; relevance of incoming information has to be evaluated separately for each threat component. The obtained information (evidence) represents an input to an argumentation system and again has to be evaluated for relevance. Relevant evidence is defined by Walton [55] as "evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence."

Quantification of the level of relevance traditionally is based on the following definition [61]:

"On the basis of prior evidence e, a hypothesis h is considered, and the change in the likelihood of h due to additional evidence i is examined. If the likelihood of h is changed by the addition of i to e, i is said to be relevant to h on the evidence e; otherwise it is irrelevant. In particular, if the likelihood of h is increased due to the addition of i to e, i is said to be positively relevant to h; if the likelihood is decreased, i is said to be negatively relevant."

Usually *Relevance Analysis* processes qualifying relevance are based on two methods [62]: the *Probability Covariance* and the *Mutual Information*. They are evaluated in [61], the authors of which discussed their drawbacks. Namely, while the probability covariance $R(X,Y) = E((X - E(X))(Y - E(Y))$ can be used to state whether two random variables X and Y are positively relevant if $R(X,Y) = 1$ or negatively relevant. This approach however fails to state whether they are relevant or independent when $R(X,Y) = 0$.

The main problem with the mutual information based relevance $R(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$ is that it is not capable to define the relevance degree but only say that the information components X and Y are relevant.

The authors of [61] introduced a different measure to represent relevance between two sentences:

$$R(X,Y)=\frac{H(X)}{\sqrt{\sum_x p(x)^2}\sqrt{\sum_y (\log p(y))^2}}-\frac{\bar{H}(X\,|\,Y=y_k)}{\sqrt{\sum_x p(x)^2}\sqrt{\sum_y p(y)^2}},$$

where $H(X)=-\sum_x p(x)\log p(x)$, $\bar{H}(X\,|\,Y=y_k)=-\sum_x p(x\,|\,Y=y_k)\log p(x\,|\,Y=y_k)$

$Y\in\{y_1,...,y_m\}$ and $X\in\{x_1,...,x_n\}$ are possible events Y and X, respectively, for example, two sentences to be compared. This relevance measure has advantages as compared with the *Probability Covariance* and the *Mutual Information* since it produces a relevance degree between 0 and 1, and does not require the conditional probability density. At the same time this measure as well as the *Probability Covariance* and the *Mutual Information* requires knowledge of probability distributions, which in turn requires a large library of sentences.

Relevance of arguments corresponds to the decision to enter arguments in an argumentation process and is based on the degree to which the premise supports the conclusion. Evaluation of argument relevance has to be done based on semantic similarity. It should be based not on *information relevance* but rather on *causal relevance*: "X is causally relevant to Y in context Z", which can be interpret to mean "Changing X will affect Y once Z is held constant" [60]. Causal relevance exists if the relations *if-then, coordinate relationship, successive order relationships*, *etc*. exist. In application to relevance of arguments, causal relevance means "How much does the premise influence the conclusion?" Inferring casual relevance between the premise and conclusion requires many grammatical rules obtained from the analysis of multiple domain specific corpora.

Human understanding of relevance is more than the result of natural language processing or matching algorithms. Humans can find other relevant information of arguments that is not detected by a system for a variety of reasons [53]. Humans derive relevant arguments and provide a level of relevance by using their expertise, prior experience, ideas and clues so human-based relevance is subjective. At the same time relevance obtained by automatic processes can also be considered subjective since the result depends of the selected method.

Based on [53, 56, 57, 58] where relevance is considered in the field of information retrieval, we summarize different types of relevance existing in an argumentation based human-machine systems:

- Topic or Subject relevance: Information relevance of the domain of interest expressed as key words, phrases and their parts and a stream of incoming information from multiple sources. "Aboutness is the criterion by which topicality is inferred."
- *Argumentation system relevance*: Causal and information relevance between sentences, the premises and conclusions, hypotheses, arguments or parts of arguments, stories or pieces of stories and information or information objects in the system (incoming

information, the results of interim processes, databases, etc.) as retrieved or as failed to be retrieved, by a given procedure or algorithm. Each system element has specific purpose and implementation and the relevance definition methods depend on this.
- "Cognitive relevance: Relevance of the cognitive state of knowledge of a user and information or information objects. Cognitive correspondence, "informativeness," novelty, information quality, and the like are criteria by which cognitive relevance is inferred." The literature presents multiple examples of such cognitive or "physiological (psychological perhaps?) relevance" [53]. For example, in [59, 60] the results of experiments in various areas of medicine are presented, which showed "causal connections between previously unrelated phenomena to derive relevance relations where none existed before; these relations were derived from literature and later confirmed in clinical testing." Cognitive relevance describes relations between information and the user's cognitive state. Since the ability to derive relevant arguments strongly depends on expertise, the symbiosis of human analyst and automatics system is required to improve relevance evaluation.
- Affective relevance: Relevance of the intents, goals, emotions, and motivations of a user, and information user receive.

8.1.2 Uncertainty Representation and Management

Data and information (acquired by physical or human sensors), arguments detected/produced by analysts; information obtained from intra-system processes such as argument detection and construction, are imperfect (e.g. uncertain and imprecise). Imperfection is the result of "partial knowledge of the true value of the data" and arises from either a lack of information or imperfection of both formal and cognitive models [63, 64], human error, and malicious intent. For example, Imperfection can be represented and managed within different theories such as probability, Bayesian probability, belief, interval probability, possibility and fuzzy set theories, *conflict* by belief and possibility theories. Selection of one of these theories depends on context, existence of prior probability, type of information (soft, hard, or both), whether the hypotheses about the state of environment under consideration are exhaustive, etc. For example, probability theory can be used to deal with repeatable experiments producing objective relative frequencies while belief, possibility, or fuzzy theories are used to represent credibility (believability) of information which is not completely trustworthy.

The Transferable Belief Model (TBM) [65] is suggested here as the one of the most appropriate for the uncertain dynamic threat environment. The TBM is a two-level model, in which quantified beliefs in hypotheses about an object or state of the environment are represented and combined at the *credal* level while decisions are made based on probabilities obtained from the combined belief by the *pignistic* transformation at the *pignistic level*. Dempster–Shafer beliefs [66], probability, and possibility [67] distributions can be expressed as belief structures represented in the framework of the TBM allowing representing both soft and hard information [68]. Beliefs are sub-additive, which permits for numerically expressing uncertainty and ignorance. Within the TBM, the unnormalized Dempster's rule can combine basic belief masses based on multiple pieces of evidence, and allow for incorporation of belief reliability.

Moreover, the TBM works under the open world assumption, i.e., it does not assume that the set of hypotheses under consideration is exhaustive. It also permits to represent conflict. These properties of the TBM have been successfully exploited in information fusion in general and in the threat context specifically (see for example [69-72]).

Formally let Θ be a set of atomic hypotheses about the state of the environment or an identity of an object: $\Theta = \{\theta_1, ..., \theta_k\}$. Let 2^Θ denote the power set. A function m is called a basic belief assignment (bba) if:

$$m : 2^\Theta \to [0,1], \quad \sum_{A \subseteq \Theta} m(A) = 1. \qquad (1)$$

In the majority of belief models $m(\varnothing)$ (uncommitted belief) is defined as zero (closed world assumption) while the TBM is the only belief model, in which uncommitted belief can be non-zero. The function, *Bel* is derived from the basic belief assignment:

$$Bel(A) = \frac{1}{1 - m(\varnothing)} \sum_{B \subseteq A} m(A). \qquad (2)$$

There is one to one correspondence between basic belief assignments and beliefs defined by (2).

If m_1 and m_2 are basic belief assignments defined on Θ, they can be combined at the *credal level* with TBM by conjunctive combination or unnormalized Dempster's rule, defined as:

$$m^\Theta(A) = \sum_{B \cap D = A} m_1(B) m_2(D), \; \forall A \subseteq \Theta \qquad (3)$$

There are special types of belief functions, which are especially suitable for representing evidence coming from multiple sources, i.e., simple and separable support functions. *Bel* is a simple support function with focus *A* with support *s*, if $\exists A \subseteq \Theta$ such that $Bel(B) = s \neq 0$ if $A \subseteq B$, $B \neq 0$, and $Bel(B) = 0$ otherwise. Separable support function is a combination of simple support functions. If *Bel* is a simple support function with focus $A \neq \Theta$, then:

$$m(A) = s, \; m(\Theta) = 1 - s, \text{and } m = 0 \text{ otherwise.} \qquad (4)$$

Belief combination at the *credal level* in the TBM follows by decision making at the *pignistic level* by using *pignistic probability*:

$$BetP^\Theta(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m^\Theta(B)}{1 - m(\varnothing)}, \quad \forall A \subseteq \Theta \;, \qquad (6)$$

where $|A|$ is the number of elements of Θ in A.

The TBM allows for declining with variable reliability of sources by considering "discount rules," which are the methods of transforming credibility of each source represented by basic belief assignments to account for their reliability and then use these transformed beliefs in the Dempster's rule of combination. In general these methods use reliability coefficients to redistribute the degree of support for different hypotheses based on reliability of beliefs into these hypotheses.

There are several ways of building discounted basic probability assignments ($\overline{m}^{disc}$). One of them is defined for simple support functions m with atomic hypothesis θ_k as a focal element to "discount" beliefs into this hypothesis by R_k .

In this case for each source *l* we will have:

$$m_k^{disc}(A) = R_k m_k(A),\ \forall A \subset \Theta, \\ m_k^{disc}(\Theta) = 1 - R_k + R_k \cdot m(\Theta) \tag{7}$$

As it was mentioned above, one of the attractive properties of the TBM is the fact that the basic belief masses can be successfully used to represent Dempster– Shafer beliefs, and probability and possibility distributions [68] allowing for fusion of multiple uncertainty representation. This property as applied to threat assessment in the framework of belie-based argumentation was discussed in [73].

Thus a probability distribution, which can characterise output of certain sensors can be represented as a Bayesian belief structure [66] $m^{pr} = P$, in which focal elements, i.e. subsets of the frame of discernment in which basic belief assignments is not zero, are singletons. Possibility distributions usually come from linguistic propositions representing observers' confidence in the evidence they supply (e.g., not sure, sure, absolutely sure). This confidence in turn represents confidence in arguments in our case, and can be used for dealing with uncertainty and imprecision of soft information. Let C be a confidence of an observer, which takes its value in the space X , and $\Pi : X \to [0,1]$ then for each $x \in X$ possibility distribution $\pi(x)$ means the possibility that x has value C . A possibility distribution can be viewed as the membership function of the fuzzy set of variable *x*.

In the framework of belief functions, a possibility distribution Π can be represented by a belief structure with nested focal elements. Let us assume that the elements in X are indexed in decreasing order of possibility, i.e. $\pi_i \geq \pi_j$ if $i < j$. Then we can represent the possibility distribution by using a Dempster-Shafer belief structure m_{poss} with focal elements $F_j,\ j = \overline{1, J}$:

$$m_{pos}^k(F_j) = \pi_j^k - \pi_{j+1}^k \text{ with } \pi_{J+1}^k = 0 \text{ by convention.} \tag{5}$$

After both probability and possibility distributions are now expressed as belief structures represented in the framework of the TBM, they can be combined with beliefs assigned to evidences from set S_1^k with the Dempster rule to obtain a belief structure *m* over hypotheses *H*, which will be used for computation of pignistic probability and decision making.

8.1.3 Trust and Reliability

There are multiple definitions of trust and reliability and there is no consensus among theorists on how to define trust [74]. At the same time most approaches rely on some version or another of the conception proposed in [75], for which trust is “a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another” [76]. As it is stated in [77] “trust must be viewed as a layered notion in its basic meaning of subjective trust, trust is “a belief, attitude, or expectation concerning the likelihood that the actions or outcomes of another individual, group or organization will be acceptable or will serve the actor’s interests” [78]. Having this in mind we define trust as a subjective information quality characteristic which can be defined as a *subjective* level of belief of an agent (either human or computational) that information he is using is sufficiently *reliable* (objective quality characteristics) and can be admitted in the system.. Trust in information/arguments should be considered in relation to the user goals and functions and in a particular context. It has to be evaluated at various steps of an argumentation system, i.e. when information enters the system, at each inter-process step, at the time when the argumentation result is presented to the analyst. Utilization of the notion of trust and its managements “aims at supporting decision making in the presence of unknown, uncontrollable and possibly harmful entities” [79].

As it can be seen in Figure 10 that trust in information as a subjective quality characteristic can be defined by reliability of the information source (physical, human sensors or processing results) and information presentation (trust in automation), which then alone or in combination of other meta-data characteristics will serve for establishing the level of trust [74]. There are several types of reliability (trust) to be considered: [52, 80, 81]

- Reliability measuring historical correctness of a source such as historical correctness of intelligence analyst or fusion result (experience or reputation-based reliability).
- Credential-based reliability defined by interaction with other sources.
- Reliability from expert opinion
- Reliability defined by the level of training
- Reliability as a second level of uncertainty, which measures reliability of the level of belief (credibility) assigned to the piece of information/argument by a human, or obtained as the result of automatic processing. Consideration of this type of reliability is especially important for belief combination in general and belief in arguments in particular, since the belief combination methods assume that the sources are equally reliable and ignoring belief reliability can lead to a skewed fusion result. This type of reliability is usually represented by reliability coefficients $\alpha \in [0,1]$.

The last definition is introduced since, for example, a human source can be truthful or does not have malicious intent but the level of belief he assigned to a certain statement can be wrong. Ideally, reliability of a source has to be evaluated by combining all these reliability characteristics. It is important to notice that credible information may not be reliable and reliable information may not be credible. It can be seem from the definition of different types of reliability that we can have either direct or indirect reliability.

Source reliability can be incorporated in the argument or input information combination by utilizing so called “quality control, which can include [52]:

- Eliminating messages, physical sensor processing results, or arguments of insufficient reliability from consideration. The level of reliability to be considered insufficient (source is not trusted) depends of the user’s needs while a user can be either an analyst or an automatic process.
- Incorporating reliability into models and processing by modifying the fusion processes.
- Modifying beliefs into information/arguments by compensating for its quality before processing or presenting to the users
- Delaying transmission of information to the next processing level or to decision makers until it has matured as a result of additional observations and/or computations improving its reliability (any-time decision making)
- Combination of strategies mentioned above.

Two types of information used in argumentation (hard and soft) have to be processes separately for building argumentation and computing their credibility to address the problem of different belief representation. Hard data is obtained as the processing result of physical sensors (acoustic, imaging, etc.) as well as the result of automatic processes such as automatic argument extraction. Information flow of the process defining reliability of hard incoming information at time *t* observed by a single sensor (direct reliability) is represented in Figure 12. As it is shown in Figure 12, reliability of hard information is defined by applicability of a sensor in a specific context obtained from domain knowledge; statistical information corresponding to sensor performance and applicability of the sensor model in the context under consideration.

A similar information flow can be considered for obtaining direct reliability of an argument is by assessing reliability of the arguments mining process. At the same time reliability of the argument mining process requires taking into account reliability of the sources of the information used as input into this process.

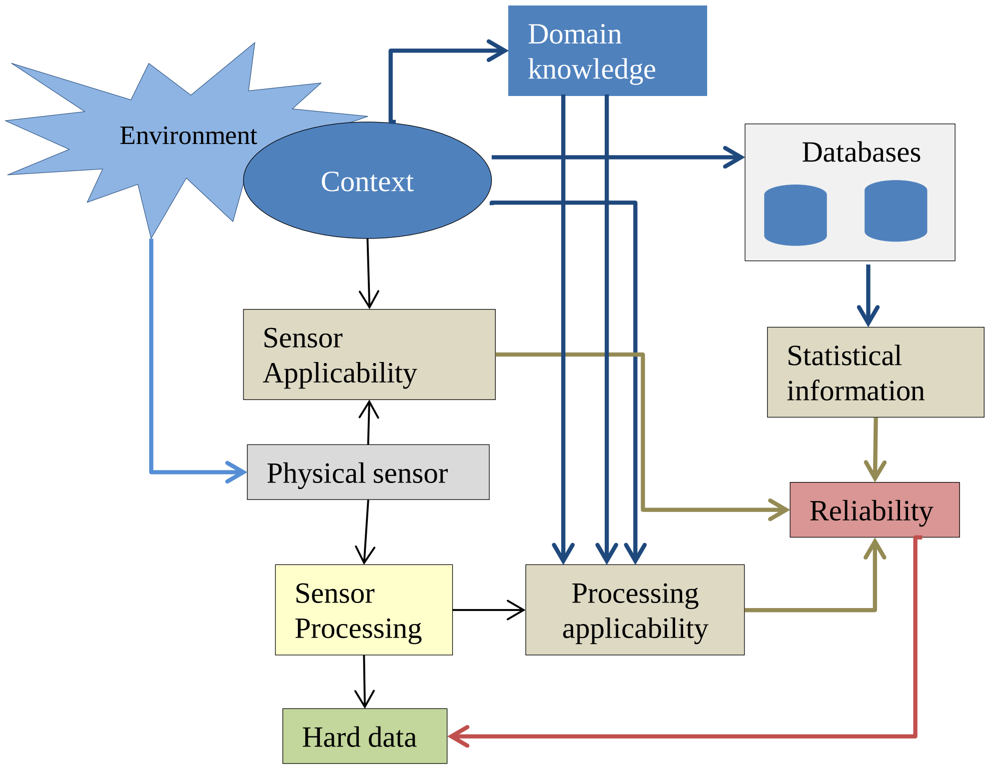


Figure 12. Hard information reliability (time t)

Since we consider dynamic situation when information/ obtained at time *t* has to be fused with relevant and reliable information obtained at time t+1. It is possible that several copies of the same message can enter the system from a sensor at different time, which may erroneously increase it reliability. Thus it is necessary to follow the history of these messages (provenance or pedigree) to avoid this problem. The issues of provenance will be discussed later in this subsection.

While direct reliability of hard data can be obtained from domain knowledge and statistical information based on the previous experience/experiments, defining reliability of soft data is a more difficult problem. For example, sources of soft information can be unreliable if they do not have incentives to tell the truth or enough knowledge about the context, in which observations are made. Another problem is that the soft information is rarely characterized by direct reliability since in many cases it comes from a network of agents with variable reliability, for example, from social networks.

One source of soft information utilized in argumentation systems is an analyst who processes incoming information presented to him and assigns a level of reliability to arguments based on reliability and timeliness of incoming information, and his level of trust in information presented to him. In addition reliability of such argument depends on reliability of the analyst judgment (reliability of argument from expert judgement). Expert is can be defined as ‘‘someone who is epistemically responsible for a particular domain of knowledge’’ and experts do not know something through intellectual trust in others, but knows something ‘‘for himself’’

[82]. Expert has expertise in a particular area makes his assertions reliable—more likely to be true than false [83]. At the same time, in order to assume that expert opinion completely reliable, it is important to take into account his characteristics (education, experience, prior and tacit knowledge, history of judgements) and understanding of context.

There are several issues to consider in modeling indirect trust such as: (see, e.g. [83-90]):

- how sources such as social media can be manipulated
- how one should revise one's notions of trust based on the past actions of individuals
- which of several competing sources of conflicting information one should trust
- how to take into account reliability of each individual.
- what are the methods of propagating reliability through the "reliability network?

To address the majority of these issues, it is necessary to consider their provenance (pedigree). Provenance as defined in [92] is information about entities, activities, and people involved in producing a piece of data or thing to be used to form assessments about their quality, reliability or trustworthiness. Establishing the reliability of information used for detecting and constructing arguments is imperative for correctly reasoning about them, and making decisions about what is going on. Provenance defines the origins of information and how and by whom this information is interpreted before entering the system. Provenance is used to construct a trust network, in which nodes represent information sources and links the level of trust between a pair of sources.

A general provenance models available on the web is PROV-DM [92] represents "a generic data model for provenance that allows domain and application specific representations of provenance to be translated into a data model and *interchanged* between systems. PROV-DM, a conceptual model allowing domain and application specific representations of provenance to be translated into such a data model as well as *interchanged* between systems. Thus, heterogeneous systems can export their native provenance into such a core data model, process it, and reason over it. For example, in [86] a model of argumentation scheme exploits PROV-DM while drawing and assessing conclusions.

Figure 13 shows an information flow in information processing for obtaining reliability of an argument constructed by an analyst. Provenance there is represented by a reliability network, in which nodes denote variable sources of information, and directed links define reliability of information transferred between them. Reasoning about reliability of each node results in reliability of incoming information presented by an analyst, who evaluates its trustworthy and defined arguments. Reliability of this argument is a combination of incoming information and reliability of the analyst.

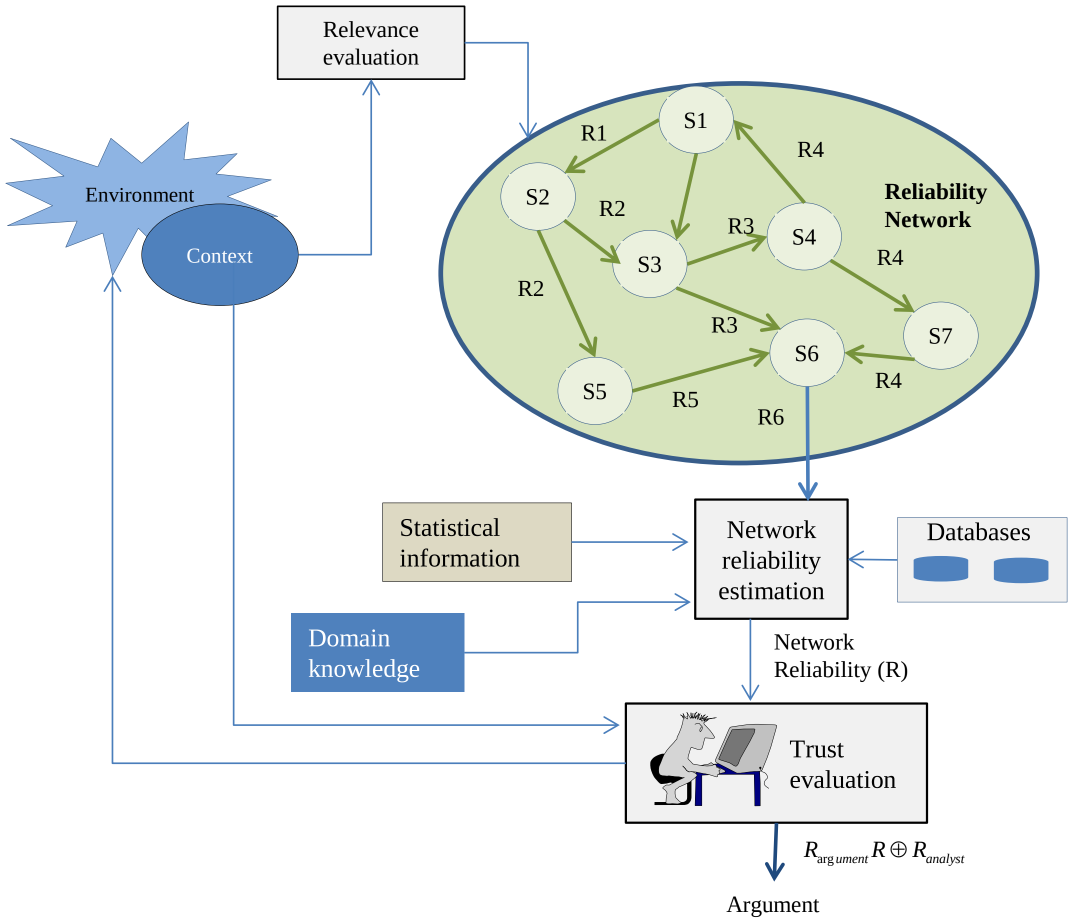


Figure 13. Information flow in the processes of obtaining reliability of an argument constructed by an analyst

Implementation of the processing shown in Figure 13 requires methods for defining relevance of information transmitted between nodes, relevance propagation involving a combination of trust in an individual node, a method of constructing the network, dealing with cycles in this network, understanding node independence, a criteria to define whether the reliability of information presented to an analyst is trustworthy enough for build an argument and assign reliability to it. A review of various models of reliability and trust evaluation is presented in [94]. If reliability is represented by probabilities or beliefs, reliability networks can be considered as belief networks and belief propagation is similar to intercausal reasoning in probabilistic networks [87,93]. In [91] a qualitative model of trust (reliability) evaluation by argumentation is offered.

9. Computational Support for Narrative Development

As described earlier, for a broad range of intelligence analysis requirements, the desired final output of analysis is a situational picture of some type. In most cases these situations are best communicated as a story or narrative description (e.g., see [21]). However, none of the system concepts and prototypes reviewed here addresses the issue of providing computational support to narrative development. In this next section, we describe our team's approach and some actual prototyping (done by Lockheed in conjunction with Virginia Tech in a separate effort).

9.1 Using Topic Modeling to Assess Story Relevance and Narrative Formation

As was remarked in particular for Section 6.3 of this report, here too we note that some elements of this section were extracted closely from the conference paper that reported the original work on Topic Modeling carried out in part by Lockheed ATL; see [95] for the original paper.

Storytelling as a data-mining concept was introduced by Kumar et.al. in [96]. Storytelling (or "connecting the dots") aims to relate seemingly disjoint objects by uncovering hidden or latent connections and finding a coherent intermediate chain of objects. This problem has been studied in a variety of contexts, such as entity networks [97], social networks [98], cellular networks [99], and document collections [100-103]. The unsupervised learning technique for storytelling called Story Chaining links related documents in a corpus to build a story or narrative arc [100]. The story chaining approach uses a real-time, flexible storytelling approach that can be used for streaming (online) data as well as for offline data. Because it is fully unsupervised, this approach does not carry the costs of competing approaches such as the need for configuration with domain knowledge or labeling of training data. As such, Story Chaining is ideal for new and frequently evolving domains. Figure 14 presents an example of a story chain generated from a corpus of news stories published in Brazil in 2013. The story chains generated from this approach can potentially tell a story about what is happening over time and across news articles by focusing on how the same people, organizations, and locations occur between documents. For this reason, story chains may be considered to be a narrative structure.

Because story chaining is an unsupervised, automated process that generates many results, there is a need to identify the story chains that contain the clearest narratives. Shahriar et.al [100] uses context overlap as a measure to produce stories that stick to one context by extracting context sentences from a document using a Naive Bayes classifier. The authors, for assessing quality, also use dispersion plots and dispersion coefficient to evaluate the overlap of contents of the documents in a chain and thereby quality. Shahaf et.al. in [102, 103] define concepts of chain coherence, coverage, and connectivity that offer more insights into the storytelling process. Our approach differs in that it learns a topic model over the corpus and tries to associate certain types of topic change across a story chain as an indicator of how clear of a narrative structure is contained within a story chain.

Topic models are probabilistic models for uncovering the underlying semantic structure of a document collection based on a hierarchical Bayesian analysis of the original texts [104]. They have been applied to a wide range of text to discover patterns of word use, or topics, across a corpus and to connect documents that share similar structure. In this way, topic models provide a way to create a structure from unstructured text in an unsupervised manner. We leverage them in our work primarily for this reason.

In our research, we have investigated the use of topic model based analytics to evaluate the clarity of the story chain narrative structure. This work proposes two different kinds of measures of assessment, representativeness and quality.

Firstly, we considered a measure of representativeness that captures how well a story chain represents the corpus from which it was generated. For example, the story chain in Figure 13 was generated from a corpus of thousands of documents published in Brazil in 2013 and it tells a clear story about the Pope visiting Brazil. The stories in the chain take place over a period of 11 days and fit well with the dominant theme of the corpus during that time period which focuses on social issues and protests. Our measure of representativeness is assessed by comparing the similarity of topics found over time in a story chain against those expressed in the corpus during the same time period. This measure assumes the corpus contains dominant topics that are desirable to understand. Our hypothesis for investigating representativeness was the idea that story chains with similar topic expression to the corpus will convey narratives that are central to the corpus.

Secondly, we considered a measure of quality in which higher quality story chains exhibit a characteristic of focusing on a small number of stable topics, rather than many interleaved or shifting topics. To evaluate this form of quality, we decomposed the measure into two contributing measures, topic persistence and topic consistency.

Topic persistence was designed to capture volatility in topic focus within a story chain. In other words, how often does the topic of a chain shift across each link in the story chain? For example, consider a story chain that has 11 articles such that there are 10 transitions in the story chain connecting one article to the next article in the chain. Topic Persistence (TP) will indicate how well topics persist between links. If most of those 10 transitions represent a change in the main topic of the article, then that story chain would have a lower TP score than a chain in which most of those 10 transitions represented no change in the main topic. In this way, if a story chain has a high TP score, then most of the links in the chain represent connections between 2 articles that are discussing the same main topic, and hence, the narrative structure is exhibiting more stable structure for a, hypothetically, better quality chain.

Topic consistency (TC) is a relative assessment of the stability of the main topic of the story chain. More specifically, TC assesses how regularly the main topic of the story chain appears as a main topic of an article within the story chain. For example, if a story chain is made up of 10 articles and has a main topic of political unrest, TC will indicate how stable that main topic of political unrest is by looking at each of the 10 contributing articles and seeing is political unrest appears as the primary topic within those 10 articles. If only 3 of those 10 articles are focused on political unrest for a TC = 3/10 or 30%, that means that most of the articles in the chain are focused on (1) different topics, and (2) a variety of different topics such that consensus did not exceed 3. Compare this to a scenario in which the story chain had 7 articles focusing on political unrest where TC = 7/10 or 70%. In this case, the topic is much more consistent throughout the chain (not necessarily consecutively) and hence, the narrative structure more centered on political unrest and, hypothetically, of better quality.

Our results indicate that using topic model based analytics to predict the quality of a narrative structure is a promising avenue of research. We found correlations between all of our analytics and the human scoring of our story chains, with particularly strong correlation to the relevance metric.

The need to build situational awareness from increasingly large sets of textual data means we must have automatic methods to construct narrative structures from text without regard to domain factors such as actors, event types, etc. The metrics presented in this paper provide a means to assess these narrative structures so that only the most useful narrative structures are transformed into narratives. In this work, we define three metrics of relevance, topic persistence and topic consistency to assess narrative structure. We specify and implement these measures with respect to a narrative structure of story chains generated by an unsupervised narrative generation technique presented in [100]. This data is processed to provide analytical evidence for the usefulness of these metrics for identifying high quality story chains.

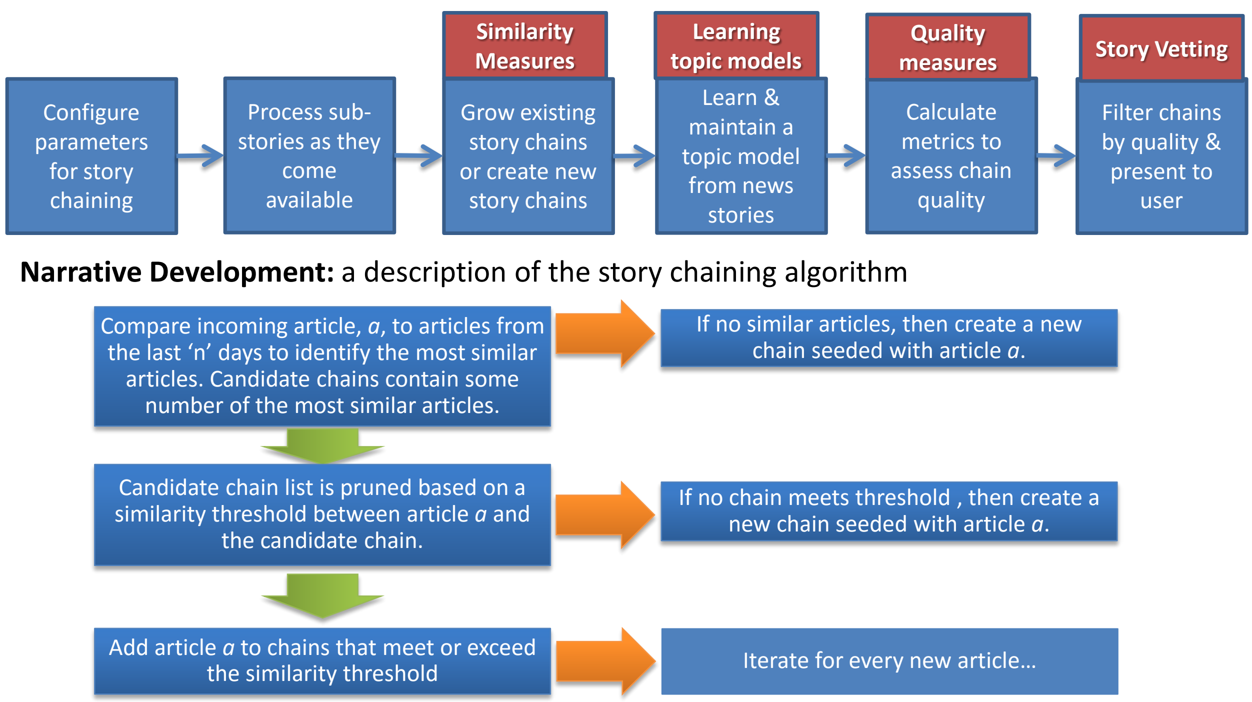


Figure 14 Overview of Topic Modeling Strategy for Narrative Development

10. A Comprehensive Look at Top-Level Functional Design

Before attempting to develop a specific functional design, we expended some effort in ranging over broad notions of a complete functional design, as a basis to frame our thinking. Here, we describe those thoughts for these high-level framework concepts.

The following is a top-level functional design of a notional prototype system that incorporates all of the research done during this effort. The goals of the system should be to make data collection easier and more relevant for the analyst. The system should work with the analyst to ask and form relevant questions and to help in finding the answers. Initially, the data collection will be collaborative with the option to automate as much as possible to offload the analyst. The analyst should be able to visualize the data in a simple and concise way and allow for multiple hypotheses to be created and explored. The ultimate goal of the system is to build out a set of structured arguments, determine COAs, and generate a set of narratives or stories of what the data represents. The collage of functions that a totally complete system should address is shown in Figure 15.

Figure 15 Comprehensive Collage of Desired Functionality

Data Collaboration

The system should allow the analyst to define goals using context and intent to automatically drive the data collection of all relevant data. Multiple hypotheses need to be created and managed as the data is fed into the system. As data is collected and organized, the missing gaps in the data can be identified and new data sources can be used to fill in the pieces.

Data Collection (Raw Data)

We expect to process both hard data and soft data. Hard data contains specific track about specific entities. Soft data contains references to these entities that need to be fused together to create a set of possible outcomes. (Need some examples of raw data.)

Data Fusion (Processed Data)

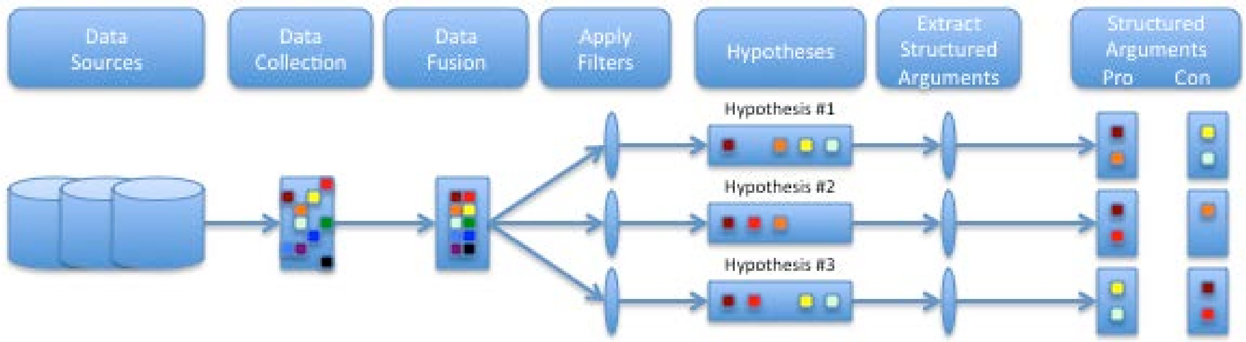
In order to manage and keep track of all the data flowing into the system, we need to be able to organize the data into searchable fused knowledge. The context of what the analyst needs to find will help to reduce the cognitive load and allow the analyst to concentrate on more important tasks. Various parameters (e.g. trust, relevance, reliability, bias, belief, rigor, source, quality, time, probabilities) will be assigned to entities within the data. This will allow the system to automatically apply fusion algorithms, create index graphs and relationships, help to align the data in time, and make it easier for the analyst to find information.

Sense Making and Decision Making

Once a process is in place to collect, store and fuse the data, the analyst and/or the system will be able to mine the data for relevant information. The analyst can perform "what if" analysis and determine how the different contexts or filters change the state of each hypothesis, template, or narrative.

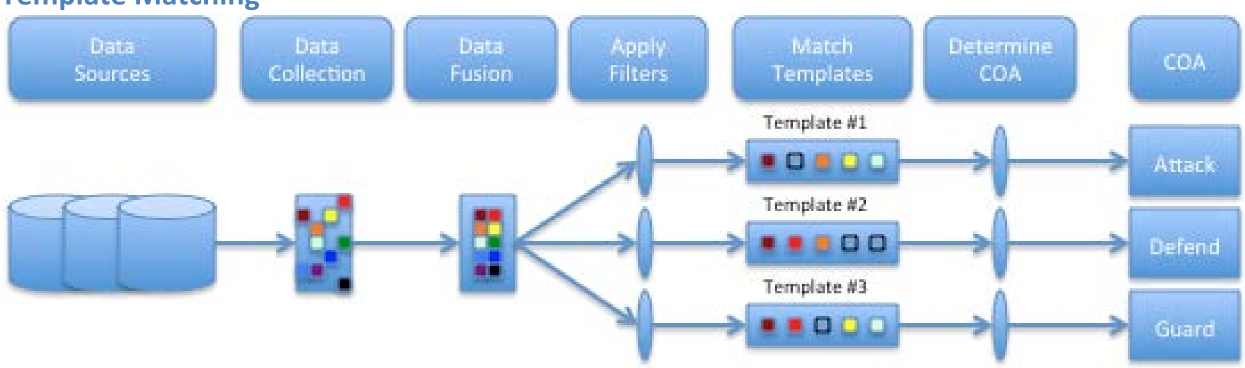
After the data has been processed and presented to the user, the analyst can then make decisions about the data that was collected. Structured Arguments can be built for each of the hypotheses to show the pros and cons of the arguments being made. COAs can be determined based on the templates applied to the data. Stories can also be created based on the narrative derived from the data.

Multiple Hypothesis Management



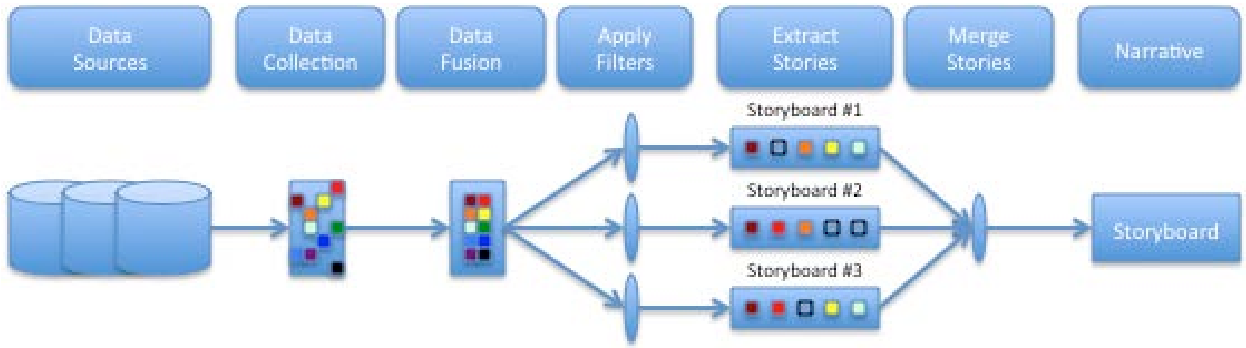
The system needs to be able to help the analyst manage and maintain a set of hypothesis about the data being collected. Filters are applied to the fused data to extract relevant data that matches each hypothesis. Structured arguments can then be extracted that helps to explain which data supports or refutes each of the hypotheses.

Template Matching



The system needs to be able to also manage a set of templates that can help guide the analyst to find gaps in the knowledge and find where to fill in the holes. Filters are applied to the fused data to match the data to each specific template. The templates could then be used to trigger when and how data is collected to generate a set of COAs.

Extract Stories



The system needs to be able to use filters to search for information with in the fused data to generate a set of storyboards. These storyboards are then merged into an overall narrative that can explain what was found in the data.

Visualization

The system should allow the analyst to visualize the data using different perspectives, generate information histograms, and validate the results as they are processed. The analyst should be able to redirect the focus of the data processing and make changes as needed to the data that is found. The system should be able to present the data in multiple perspectives to allow the analyst better insight to what is stored in the data.

11. Developing a Functional Design for an Advanced-capability Prototype

An effective approach to architecting our proposed decision-support concept requires that we assert our views of the overall reasoning process from evidence to decision-making and decision enablement. Most traditional characterizations describe decision-making (DM) as contemplative, analytic, involving nomination and evaluation of options that are weighed in some context, eventually leading to a choice of a "course of action (COA)". This model, often labeled as the "System 2" model, can be seen in most descriptions of the "Military Decision-Making Process" or MDMP as for example in published military Field Manuals such as in [105]. The literature also identifies a "System 1" or largely intuitive decision-making paradigm (IDM) that operates in conjunction with System 2 processes in what is argued to be an improved DM process model, often called the "Dual-Process Model". Most research in decision support however has focused on System 2 DM ideas since this model is quantitative and can be mathematically studied using notions of utility theory and other frameworks for mensuration. We intend however to factor the Dual-Process Model concept into our systemic design approach; the basis of this rationale cannot be elaborated here but we offer our references for the interested reader, e.g., [106, 107].

Furthermore, in our view of the System Support context for DM, we see what today are called Sensemaking processes, as lying between automated System Support capabilities such as Data Fusion processes and DM processes, in a stage wherein "final" situation assessments and understandings (in the human mind) are developed. Thus, our view of this meta-process is as a three-stage operation: System Support (SS) as an automated process that nominates algorithmically-formed situational hypotheses (such as from the combined operations of data fusion and argumentation), followed by human-computer, mixed-initiative processes for Sensemaking and symbiosis, whose narrative-type products provide the vetted situational assessments needed for decision-making. There is a substantive literature on Sensemaking, such as those previously cited [108, 109]. Our key thoughts on and rational for the meta-architecture for System Support described briefly here have been summarized in [105]. Finally, in the face of significant production pressures and rapidly proliferating data availability —and the resulting data overload deluging the professional analyst—it is increasingly easy for analysts and decision-makers to be trapped by shallow, low-rigor analysis; improvements in rigor have been previously discussed and are part of our proposed design. At the highest level, and consistent with the System Support/Fusion—Sensemaking--Decision-making interdependent processes concept, we see our initial prototype as embedded in the Sensemaking dynamic (note that this is an initial, design-in-process), as shown in Figure 16:

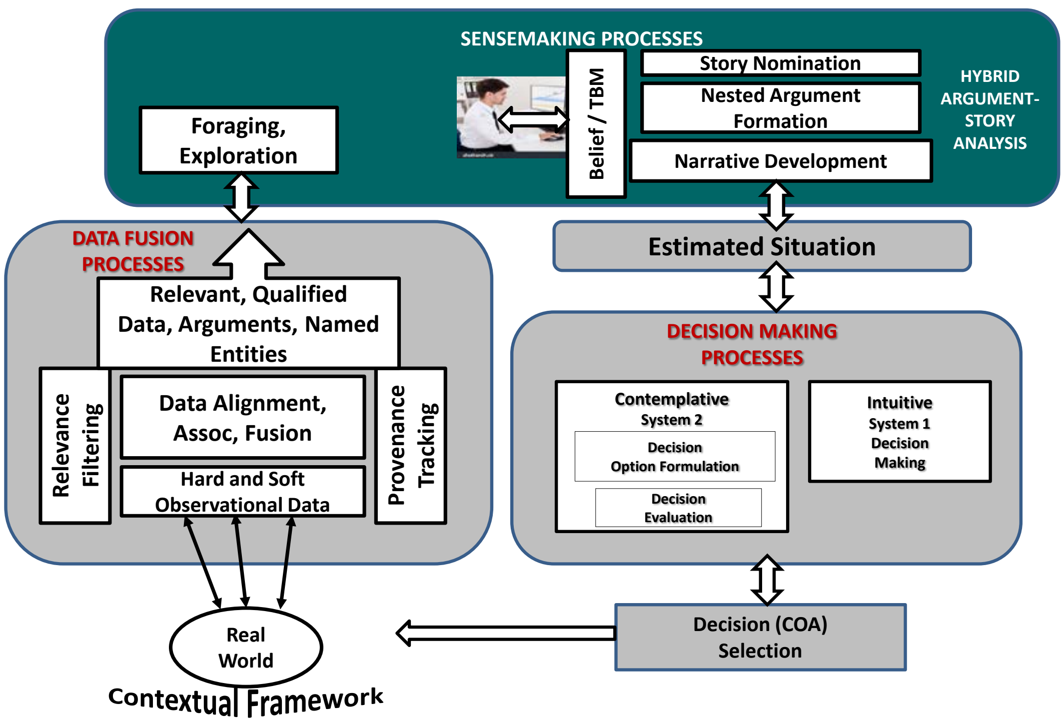


Figure 16 The Hybrid Scheme in the Context of a Meta-Architecture Involving Fusion-Sensemaking-Decision-Making (see [108])

Building on these ideas, we formed our initial functional design as shown in Figure 17. Included in this design are the specifics of the Hard-Soft data association operations that would be part of the Fusion/System Support segment in an eventual final design. The figure can be examined by starting at the bottom where notional Use Cases are also shown—these include current service-specific mission operations, Joint service operations, and a technological type thrust that examines the proposed methods as having disruptive properties:

- **Army: Operations in Megacities, Syrian Civil War**
    - **Megacity operations are an evolving new Army interest**
- **Navy: Piracy (NATO), Autonomous ISR Systems**
    - **Piracy is a continuing NATO interest, ONR has considerable interest in UAV/UXV operations**
- **Joint: Expeditionary Operations (Anti-Access Area Denial, A2AD),**
    - **Joint operations dealing with A2AD issues are an evolving widespread interest**
- **Assess Hybrid Argumentation Technology as Disruptive**
    - **And of course these proposed methods can be studied from the technological point of view as a new and disruptive capability**

For any Use Case, we envision that there would be the opportunity or need to enable both Hard and Soft data stream inputs of various types as peculiar to each of the Use Cases. Using the "Foundational" ideas of Section 8 especially in regard to forming computational support techniques for Relevance filtering and Provenance accounting, we show those two functional blocks first, operating on both data streams. (Note that there may be some preprocessing required for the Hard Data stream to frame the results into Entity-Attribute sets.) These filters ideally provide relevant and qualified data to two processes: a Natural Language Processor (NLP) and Argument Detection and Nomination (ADM) process. The functions of each of these operations are:

- NLP: extract Named Entities and associated features and attributes of those Named Entities
- ADM: detect and construct argument phrases with labeled Schemas as possible

Metadata is also available for both processing operations. The outputs of both NLP and ADM (and possible Hard Data preprocessing) are inputs to the Hard/Soft Data Association process that correlates the Entity-Attribute sets and forms the associated and reconciled fused Entity/Attribute results, i.e., the associated, fused Entity/Enriched Attribute evidential data set as shown on Figure x. This output provides a feedback to the Argument Detection processing (that contains labeled Entities) so that these identified Entities can be enriched with the associated/fused Attributes. Note that there can be possible outlier Entities here, since the ADM process is only Soft-data-based; this is a reconciliation issue yet to be determined. One idea is to engage the human analyst in the process of integrating and managing these outlier Entities. At this point, this front-end processing has ***automatically*** produced nominated arguments with associated and enriched/fused Entity/Attribute pairs—this capability is a high-priority goal of our approach as this capability has the potential to greatly reduce human cognition workload in terms of argument construction, a major issue even in the most modern prototypes we have reviewed. These nominated arguments then are vetted with analyst intervention and once vetted can provide draft input to our proposed Topic Modeling/Narrative Construction software that aids in a mixed-initiative, human-machine symbiotic process of hybrid argument/story combination. These operations will likely involve the management of competing hypotheses for which Lockheed IRAD software may also provide automated support. These operations would take advantage of Bex's theories and methods for hybrid correlation of the evidentially-grounded arguments and stories emanating both from the analyst and from the Topic Modeling story-nomination process.

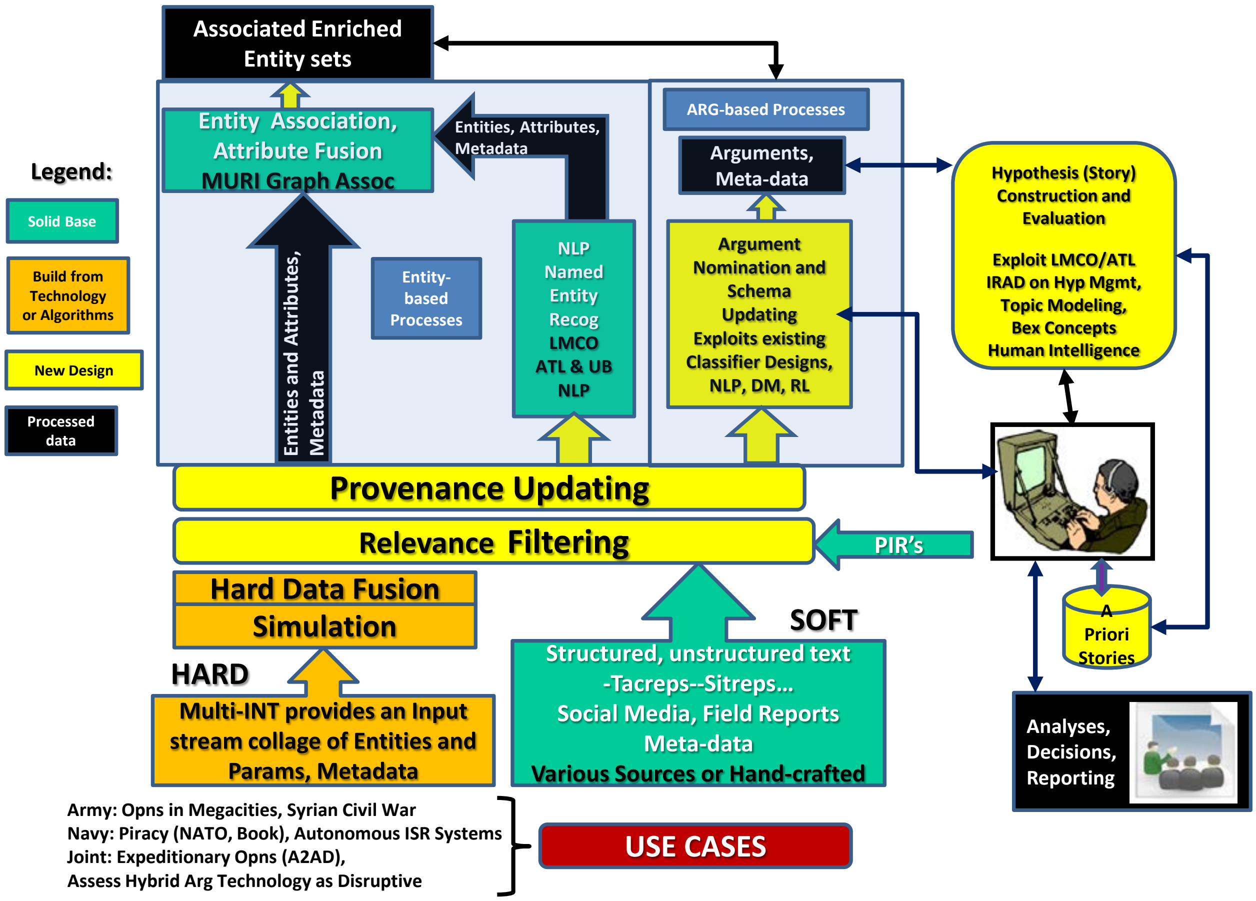


Figure 17 The Final Functional Design

This is of course an ambitious vision but is one that sets a new milestone we think for automated support to intelligence analysis. A number of details have to be worked out but the considerably advanced capabilities that a system like this can provide will move the bar forward in terms of revolutionary, disruptive automated support to intelligence analysis.

11.1 Looking Ahead: Possible Test and Evaluation Schemes

Given that our end-goal of this project was to develop initial thoughts on a functional design, it was considered necessary to explore possible strategies for Test and Evaluation (T&E) as well as possible metrics for evaluation, since the quality of any possible prototype would be measured by some appropriate T&E approach.

There are various important functions in the proposed top-level design of Fig 16. As the multisource Data Association process is considered key in any Information Fusion process, one critical aspect of a T&E approach would suggest a scheme for evaluating Hard-Soft Data Association. Here, we would suggest the approach of the MURI program that the Center for Multisource Information Fusion at the University at Buffalo developed as at least a starting approach (this is well-documented in [109, 110]); this technique was explored and tested with good success on that program.

Testing of Natural Language Processing (NLP) methods is a very broad topic but one focus for the proposed design in on Named Entity extraction, a key capability for good performance in the proposed scheme. Here too the methods employed on the prior MURI program could be applied to evaluate performance in any Use Case application; these techniques are discussed in [111].

There is not much literature on specific evaluation techniques for the various front-end argument detection/construction methods we would intend to explore, but most of these rely on some type of classification framework, and evaluation of such text extraction methods. The cited literature of Section 6, along with various survey papers on classifier evaluation form an adequate starting point for developing an evaluation approach.

Evaluating the quality of argument constructs is an area where there is considerable literature. There are various websites on this topic (e.g., http://www.csuchico.edu/~egampel/students/evaluating.html)) and a wide variety of papers that address this topic (e.g., [112]). Much of the literature discusses notions of argument strength, different for deductive, inductive, and abductive arguments and introduce related ideas on validity of premises and other issues. This literature is helpful toward test planning but we prefer Dahl's ideas on the notion of argument persuasiveness that in turn relates to ideas on "explanatory coherence" as a technique for evaluating the persuasiveness of arguments; see [113-115].

Of course, the best evaluation approach would reveal the impacts of these combined technologies on mission-based analysis effectiveness; however, since the proposed design and suggested methods are, in our opinion, still at the formative stage, much testing and evaluation would have to be done to first establish technological credibility before mission effectiveness assessments could (or should) be carried out.

12. References

[1] Llinas, J. and Rogova, G., Belief-based Argumentation in Intelligence Analysis and Decision Making, Proc of the Conf on Distributed Computing and Artificial Intelligence, Seville, Spain, June 2016

[2] Analysts Notebook can be retrieved at:
http://www03.ibm.com/software/products/en/analysts-notebook
[3] Andrews, C. and North, C., "Analyst's Workspace: An Embodied Sensemaking Environment For Large, High-Resolution Displays", Proc. 2012 IEEE Conference on Visual Analytics Science and Technology (VAST), Seattle, WA, Oct., 2012
[4] Sentinel Visualizer can be retrieved at: http://www.fmsasg.com/
[5] Palantir information can be retrieved at: http://www.palantir.com/
[6] Bier, E.A., Ishak, E.W., and Chi, E., "Entity Workspace: An Evidence File That Aids Memory, Inference, and Reading", In ISI, San Diego, CA, 2006, pp. 466-472.
[7] Stasko, J., Gorg, C., Liu, Z., and Singhal, K., "Jigsaw: Supporting Investigative Analysis through Interactive Visualization", Proc. 2007 IEEE Conference on Visual Analytics Science and Technology (VAST), Sacramento, CA, Oct., 2007
[8] Pirolli, P. and Card, S., The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis, In Proceedings of 2005 International Conference on Intelligence Analysis (McLean, VA, USA, May, 2005). http://cryptome.org/intel-2005.htm
[9] Klein, G., et al, Making Sense of Sensemaking 2: A Macrocognitive Model, IEEE Intelligent Systems, Volume:21 , Issue: 5, 2006
[10] Llinas, J., A Survey of Automated Methods for Sensemaking Support, Proc of the SPIE Conf on Next-Generation Analyst, Baltimore, MD, April 2014
[11] (Licklider, J. C. R. 1960, 'Man-Computer Symbiosis', IRE Transaction on Human in Electronics, volume HFE-1, pp. 4 – 11.)
[12] Zackova, 2011 MAN-COMPUTER SYMBIOSIS AS A WAY OF HUMAN COGNITIVE ENHANCEMENT, Masaryk University Journal of Law and Technology, Vol 5.1, 2011
[13] Engelbart, D. C. 1962, Augmenting Human Intellect: A Conceptual Framework, prepared for: Director of Information sciences Air Force Office of Scientific Research, Washington 25, D.C.
[14] Roy, D., 10x—Human-Machine Symbiosis, BT Technology Jl, 22, 4, Oct 2004.
[15] Quinn, AJ and Benderson, BB, Human Computation: A Survey and Taxonomy of a Growing Field, CHI 2011, May 7–12, 2011, Vancouver, BC, Canada
[16] von Ahn, L., Human Computation, Doctoral Thesis, UMI Order Number: AAI3205378, CMU, 2005
[17] Norman, D., The Design of Everyday Things, Basic Books, Philadelphia, 2013
[18] Norman, D., Human-centered design considered harmful, Interactions, Volume 12 Issue 4, July + August 2005
[19] Llinas, J., Nagi, R., Hall, D.L., and Lavery, J., "A Multi-Disciplinary University Research Initiative in Hard and Soft Information Fusion: Overview, Research Strategies and Initial Results", Proc. of the Intl Info Fusion Conf., Edinburgh, UK, July, 2010
[20] K. Date, G. A. Gross, S. Khopkar, R. Nagi, K. Sambhoos., "Data association and graph analytical processing of hard and soft intelligence data", Proceedings of the 16th International Conference on Information Fusion (Fusion 2013), Istanbul, Turkey, 09-12 July 2013.
[21] Schum, D., NARRATIVES IN INTELLIGENCE ANALYSIS: NECESSARY BUT OFTEN DANGEROUS, UNIVERSITY COLLEGE LONDON STUDIES OF EVIDENCE SCIENCE , June 2005
[22] Smets, P., The transferable belief model, Artificial Intelligence,Volume 66, Issue 2, April 1994, Pages 191-23

[23] Bex, F., et al, A hybrid formal theory of arguments, stories and criminal evidence, Artificial Intelligence and Law, July 2010
[24] Bex, F., et al, Sense-making software for crime investigation: how to combine stories and arguments?, Law, Probability and Risk (2007) 6
[25] Bex, F. and Verheij, B., Story Schemes for Argumentation about the Facts of a Crime, Computational Models of Narrative: Papers from the AAAI Fall Symposium Oct 2010
[26] Moen, M., et al, Automatic Detection of Arguments in Legal Texts, Proceeding
ICAIL '07 Proceedings of the 11th international conference on Artificial intelligence and law, 2007
[27] J. Allen. Natural Language Understanding (2nd ed.). Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA, 1995
[28] Mochales-Palau, R. and Moens, M., Study on Sentence Relations in the Automatic Detection of Argumentation in Legal Cases, Proceedings of the 2007 Twentieth Annual Conference on Legal Knowledge and Information Systems: JURIX 2007
[29] Walton, D., Reed, C., and Macagno.F., Argumentation Schemes. Cambridge University Press, 2008.
[30]Mochales, R. and Moens, M., Study on the Structure of Argumentation in Case Law, Proceedings of the 2008 Twenty-First Annual Conference on Legal Knowledge and Information Systems: JURIX 2008
[31] Feng,V.W. and Hirst, G., Classifying Arguments by Scheme, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 987–996, Portland, Oregon, June 19-24, 2011
[32] Moens, M., Argumentation Mining: Where are we now, where do we want to be and how do we get there?, FIRE '13 Proceedings of the 5th 2013 Forum on Information Retrieval Evaluation
[33] Dan Roth and Wen-tau Yih. Integer linear programming inference for conditional random fields. In Proceedings of the 22nd International Conference on Machine Learning, ICML '05, pages 736{743, New York, NY, USA, 2005. ACM.
[34] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. Large margin methods for structured and interdependent output variables. Journal of Machine Learning Research, 6:1453{1484,
December 2005
[35] Walton. D. and Gordon, T.F., THE CARNEADES MODEL OF ARGUMENT INVENTION, Pragmatics & Cognition, 20(1), 2012
[36] Gordon, T.F. and Walton, D., The Carneades Argumentation Framework – Using Presumptions and Exceptions to Model
Critical Questions
[37] Hastings, A..C., A Reformulation of the Modes of Reasoning in Argumentation, Ph.D. dissertation, Northwestern University, Evanston, Illinois, 1963.
[38] van den Braack, S.W., et al, A critical review of argument visualization tools: do users become better reasoners?, ECAI-06 CMNA Workshop, 2006
[39] Twardy, C.: Argument maps improve critical thinking. Teaching Philosophy 27 (2):95--116 (2004)

[40] Mani, I. and Klein, G.L., Evaluating Intelligence Analysis Arguments in Open-ended Situations, Proc of the Intl Conf on Intelligence Analysis, McLean Va, 2005
[41] Gordon, T.F., Computational Dialectics Tools – Argumentation Support Systems
[42] Suthers, D., et al,, 'Belvedere: Engaging students in critical discussion of science and public policy issues', in AI-Ed 95, the 7th World Conference on Artificial Intelligence in Education, pp. 266–273, (1995).
[43] Reed, C. and Rowe, G., ARAUCARIA: SOFTWARE FOR ARGUMENT ANALYSIS, DIAGRAMMING AND REPRESENTATION
[44] van Gelder, T., The rationale for Rationale,in Law, Probability and Risk (2007)
[45] Van den Braack, S. W., et al, AVERs: an argument visualization tool for representing stories about evidence, Proceedings of the 11th international conference on Artificial intelligence and law, Stanford, CA, 2007
[46] Van den Braack, S. W., Sensemaking software for crime analysis, Dissertation, Univ of Utrecht, Holland, 2010
[47] F. Bex, S. van den Braak, H. van Oostendorp, H. Prakken, B. Verheij, and G. Vreeswijk, "Sense-making software for crime investigation: how to combine stories and arguments?," *Law, Probability and Risk*, vol. 6, iss. 1-4, pp. 145-168, 2007.
[48] CISpaces, towards a system for online collaborative argument construction and debate
[49] Standard 8402, 3. I (1986) International organization of standards
[50] J. A. O'Brien, G. Marakas, Introduction to Information Systems, McGraw-Hill/Irwin, 2005
[51] J, Juran, A, Godfrey, Juran's quality handbook, 5th edn. McGraw-Hill, New York, 1988
[52] G. Rogova, Information Quality in Information Fusion and Decision Making with Applications to Crisis Management, in Fusion methodology in crisis management: higher level fusion and decision making, (Rogova &Scott, eds), 65-86, Springer, 2016.
[53] T. Saracevic, Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. Part II, Journal of the American Society for Information Science and Technology, 58 (3) 1915-1933, 2006.
[54] H. D. White, Relevance theory and citations, Journal of Pragmatics, 43(14), 3345–3361.
[55] D. Walton, Relevance in argumentation, 2003
[56] T. Saracevic, Relevance reconsidered '96. In Information Science: Integrationin Perspective. Proceedings of Second International Conference on Conceptions of Library and Information Science (CoLIS 1996) (P. Ingewersen and N. O. Pors, eds.), pp. 201–218. The Royal School of Librarianship, Copenhagen, 1996.
[57] P. Ingwersen, Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. Journal of Documentation 52(1), 3–50, 1996.
[58] P. Borlund, P. The concept of relevance in IR. Journal of the American Society for Information Science and Technology 54(10), 913–925, 2003.
[59] D.R, Swanson, N.R., Smalheiser, N. R. (1997). An interactive system for finding complementary literatures: A stimulus to scientific discovery. Artificial Intelligence, 91(2), 183–203, 1997.
[60] D. Galles, J. Pearl, Axioms of causal relevance, Artificial Intelligence, 97, 9-43, 1997.
[61] P. Gardenfors. On the logic of relevance. Synthese, 37(3):351–367, March 1978.

[62] Mao-Sheng Zhong, Lei Liu, and Ru-Zhan Lu, A New Method of Relevance Measure and Its Applications, in Proc: The IEEE Sixth International Conference on Advanced Language Processing and Web Information Technology, 595-600, 2007.
[63] Ph. Smets, Imperfect information: imprecision – uncertainty, A. Motro and Ph. Smets (editors). Uncertainty Management in Information Systems: From Needs to Solutions. pp. 225-254, Kluwer, 1997.
[64] A-L Jousselme, P Maupin, E. Bosse, Uncertainty in a situation analysis perspective, in: Proc. of the FUSION'2003-6th Conference on Multisource- Information Fusion, pp. 1207-1214, 2003.
[65] P. Smets, Data Fusion in the Transferable Belief Model, In: Proc. of the FUSION'2000-Third Conference on Multisource- Multisensor Information Fusion, 2002, pp. 21-33.
[66] G. Shafer, A Mathematica theory of evidence, G. Shafer, Princeton, NJ: Princeton, Univ. Press, 1976.
[67] D.Dubois and H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty. New York: Plenum, 1988.
[68] R. Yager, Conditional Approach to Possibility-Probability Fusion, IEEE Transactions on Fuzzy Systems, VOL. 20, No. 1, pp 46-55, 2012.
[69] F. Delmotte, Ph.Smets, Target Identification Based on the Transferable Belief Model Interpretation of Dempster-Shafer Model., IEEE Trans. Syst., Man, Cybern. A: 34, pp. 457-471, 2004.
[70] G. Rogova, P. Scott, C. Lollett, R. Mudiyanur, Reasoning about situations in the early post-disaster response environment, in: Proc. of the FUSION'2006-9th Conference on Multisource Information Fusion, 2006.
[71] G. Rogova, M. Hadrazagic, M-O. St-Hilaire, M. Florea, P. Valin, Context-Based Information Quality for Sequential Decision Making, in Proc. of the 2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2013.
[72] G. Rogova, Adaptive Real-Time Threat Assessment under Uncertainty and Conflict, in Proc: 4th IEEE Conference on Cognitive Methods in Situation Awareness and Decision Support, San Antonio, TX, 2014.
[73] G. Rogova, J. Llinas, R. Yager, Integration of Argumentation and Fusion of Soft-Hard Information for Threat Assessment, in Proc of the NATO Joint Symposium on Information Fusion (Hard and Soft) For Intelligence, Surveillance & Reconnaissance (ISR).
[74] L. Bonelli, S. Felletti, F. Paglieri, From Argumentative Crisis to Critical Arguments: How to Argue in the Face of Danger, in Fusion methodology in crisis management: higher level fusion and decision making, (Rogova &Scott, eds), 365-381, Springer, 2016.
[75] D.M. Rousseau, S.B. Sitkin, RS. Burt, C. Camerer, Not so different after all: a cross discipline view of trust. Acad Manage Rev 23(3):393–404, 1998.
[76] R.C. Mayer, J.H. Davis, F.D. Schoorman, An integrative model of organizational trust. Acad. Manage Rev 20(3):709–734, 1995.
[77] C. Castelfranchi, R. Falcone, Trust theory: a socio-cognitive and computational model. Wiley, 2010.
[78] S.B. Sitkin, N.L.Roth, Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. Organ Sci 4(3):367–392, 1993.
[79] Ch. F. Chang, P. Harvey, and A.Ghose, Combining Credibility in a Source Sensitive

Argumentation System, G. Antoniou et al. (Eds.): SETN 2006, LNAI 3955, pp. 478–481, 2006.Springer-Verlag Berlin Heidelberg 2006.
[80] G. Rogova, E. Bosse, Information quality in information fusion, in: Proc. of the 13th International Conference on Information Fusion, Edinburg, Scotland, July, 2010.
[81] K. Krukow, M. Nielsen and V. Sassone, Trust Models in Ubiquitous Computing, Philosophical Transactions: Mathematical, Physical and Engineering Sciences, Vol. 366, No. 1881, From Computers to Ubiquitous Computing, by 2020 (Oct. 28, 2008), pp. 3781-3793.
[82] Ch. Huenemann, Why not to trust other philosophers. American Philosophical Quarterly 41(3), 2004,: 249–258.
[83] D. N. Walton, Appeal to expert opinion: Arguments from authority, University Park, PA: Penn State University Press, 1997.
[84] J. Lang, M. Spear, S.F. Wu, Social manipulation of online recommender systems, in Proceedings of the 2nd International Conference on Social Informatics. 2010.
[85] K. Lerman, A. Galstyan,. Analysis of social voting patterns on Digg, in Proceedings of the 1st Workshop on Online Social Networks, 2008.
[86] Y Wang, CW Hang, MP Singh, A probabilistic approach for maintaining trust based on evidence, Journal of Artificial Intelligence Research, 40 (1), 221-26.
[87] S. Parsons, K. Atkinson, Z. Li, P. McBurney, E. Sklar, M. Singh, J. Rowe, Argument schemes for reasoning about trust. Argument and Computation, 5(2-3), 160-190, 2014
[88] Dong, X. L.; Berti-Equille, L.; and Srivastava, D. 2009. Integrating conflicting data: The role of source dependence, in Proceedings of the 35th International Conference on Very Large Databases.
[89] F.Cerutti, L.Kaplan, T.Norman, N.Oren, and A.Toniolo. Subjective logic operators in trust assessment: an empirical study. Information Systems Frontiers, 17(4):743–762, 2015.
[90] A.Toniolo, F.Cerutti, N.Oren, T.J.Norman, and K.Sycara. Making informed decisions with provenance and argumentation schemes. In Proceedings of the 11th International Workshop on Argumentation in Multi-Agent Systems, 2014.
[91] H. Prade, A Qualitative Bipolar Argumentative View of Trust, in H. Prade, and V.S. Subrahmanian (Eds.): SUM 2007, LNAI 4772, pp. 268–276, 2007.
[92] PROV-DM: The PROV Data Model, https://www.w3.org/TR/prov-dm/
[93] D. Koller and N. Friedman. Probabilistic Graphical Models: Principles and Techniques. MIT Press, Cambridge, MA, 2009.
[94] A. Jøsang, R. Ismail, C. Boyd, A Survey of Trust and Reputation Systems for Online Service Provision, Decision Support Systems, 43(2) 2007, p.618-644.
[95] Tareq Ahram, Waldemar Karwowski, Dylan Schmorrow, Jason Schlachter, Alicia Ruvinsky, Luis Asencios Reynoso, Sathappan Muthiah, Naren Ramakrishnan, Leveraging Topic Models to Develop Metrics for Evaluating the Quality of Narrative Threads Extracted from News Stories, 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015 Procedia Manufacturing, Volume 3, 2015, Pages 4028-4035, ISSN 2351-9789, http://dx.doi.org/10.1016/j.promfg.2015.07.972.
[96] Kumar, D., Ramakrishnan, N., Helm, R. F., & Potts, M. (2008). Algorithms for storytelling. Knowledge and Data Engineering, IEEE Transactions on, 20(6), 736-751.

[97] Hossain, M. S., Butler, P., Boedihardjo, A. P., & Ramakrishnan, N. (2012). Storytelling in entity networks to support intelligence analysts. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.
[98] Faloutsos, Christos, Kevin S McCurley, and Andrew Tomkins. "Fast discovery of connection subgraphs." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 22 Aug. 2004: 118-127.
[99] Hossain, M Shahriar, Monika Akbar, and Nicholas F Polys. "Narratives in the network: interactive methods for mining cell signaling networks." Journal of Computational Biology 19.9 (2012): 1043-1059.
[100] Hossain, M Shahriar et al. "Connecting the dots between PubMed abstracts." PloS one 7.1 (2012): e29509.
[101] Shahaf, Dafna, and Carlos Guestrin. "Connecting the dots between news articles." Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining 25 Jul. 2010: 623-632.]
[102] Shahaf, Dafna, Carlos Guestrin, and Eric Horvitz. "Trains of thought: Generating information maps." Proceedings of the 21st international conference on World Wide Web 16 Apr. 2012: 899-908.
[103] Shahaf, Dafna et al. "Information cartography: creating zoomable, large-scale maps of information." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining 11 Aug. 2013: 1097-1105.
[104] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.
[105] Headquarters, Dept of Army, Army Field Manual 5-0, The Operations Process, May 2010
[106] P. Croskerry, A Universal Model of Diagnostic Reasoning, Academic Medicine, Vol 84, No 8, pp1022–8, 2009
[107] B. Djulbegovic, et al, Dual processing model of medical decision-making, BMC Medical Informatics and Decision Making, Vol. 12, Sept. 2012
[108] J. Llinas, J., Reexamining Information Fusion--Decision Making Inter-dependencies, in Proc. of the IEEE Cognitive Methods in Situation Awareness and Decision Support (CogSIMA) Conference, San Antonio, TX, March, 2014
[109] Gross, G., et al, Systemic Test and Evaluation of a Hard+Soft Information Fusion Framework; Challenges and Current Approaches Fusion 2014
[110] Date, K., et al, Test and Evaluation of Data Association Algorithms in Hard + Soft Data Fusion
[111] Shapiro, S., Tractor: Toward Deep Understanding of Short Intelligence Messages, University Seminar, available at: http://studylib.net/doc/10515245/tractor-toward-deep-understanding-of-short-intelligence-m..., 2012
[112] Corner, A. and Hahn, U., Evaluating Science Arguments: Evidence, Uncertainty, and Argument Strength, Journal of Experimental Psychology Applied 15(3):199-212 · September 2009
[113] Dahl, E., Intelligence and Surprise Attack: Failure and Success from Pearl Harbor to 9/11 and Beyond, Georgetown University Press, 2013

[114] Thagard, P., Probabilistic Networks and Explanatory Coherence, Cognitive Science Quarterly (2000) 1, 93-116
[115] Ng, H.T., and Mooney, R.J., On the Role of Coherence in Abductive Explanation, in Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI-90), pp.337-342, Boston, MA, July, 1990